PHYS 51: Electromagnetism and Optics

Connor Neely, Fall 2023

* Adapted from David J. Griffiths, *Introduction to Electrodynamics* (2013) and FA23 lectures.

1 Electrostatics

1.1 The Electrostatic Force

In mechanics, all of the quantities we worked with could be described using some combination of mass, length, and time. Starting now, though, we'll discuss a new kind of unit: charge, whose SI unit is the Coulomb (C). It is conserved in closed systems, relativistically invariant, and quantized. (The quantum is the charge q_e carried by a proton or electron, often called the elementary charge.)

Charge can be either positive or negative, and it's well know that charges with like charges repel while those with opposite charges attract. In particular, for two point charges q, q_0, the force that q exerts on q_0 is given by

$$\mathbf{F}_E(r) = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r^2}\hat{\mathbf{r}},$$

where $\mathbf{r}$ runs from q to q_0 and ϵ_0 is called the permittivity of free space. We call this relationship the electrostatic force, also known as Coulomb's law. Notice how it takes the exact same form as Newton's law of gravitation, simply with the added caveat that the particles may either attract or repel.

When we talk about gravity, the idea of a "gravitational field" often comes up to loosely describe how one massive body interacts with others around it. We can do the same with point charges—if a charge q exerts a force $\mathbf{F}_E$ on q_0, then the electric field due to q is given by

$$\mathbf{E}(r) = \frac{1}{q_0}\mathbf{F}_E(r) = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}}.$$

Notice how $\mathbf{E}$ is independent of our choice of q_0, depending only on the charge q and how far away it is. Also, electric fields obey the principle of superposition—in a space with several point charges, their net electric field is found by simply summing the individual charges' fields.

Example: Electric field due to a dipole

Electric dipoles frequently arise in scenarios involving charged objects. Consider two particles with equal and opposite charges q spaced a distance d apart; if we look at a point a distance x from the charges' midpoint, then by superposition we have

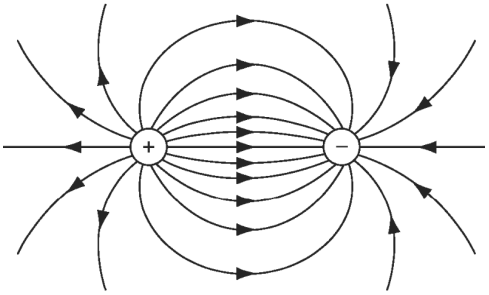
$$E_{\text{dip}} = 2 \cdot \frac{1}{4\pi\epsilon_0}\frac{q}{\mathcal{R}^2}\sin\theta,$$

where $\mathcal{R}$ is the distance from x to either of the charges and θ is the angle that the x-charge vector makes with the vertical. By geometry this gives

$$\mathbf{E}_{\text{dip}} = \frac{1}{4\pi\epsilon_0}\frac{qd}{(r^2 + (d/2)^2)^{3/2}}(-\hat{z}),$$

with $\hat{z}$ pointing from the charge midpoint toward $+q$. In general we define the dipole moment $p \equiv qd$ to describe the separation of two opposite charges. (Note that $\mathbf{E}_{\text{dip}} \propto r^{-3}$ for $r \gg d$.)

There's a nice way to visualize these fields using electric field lines. An example for the case of a dipole is provided at right. There are a few important features to note: the lines point in the direction of the electric field, never intersect, and are only drawn in the plane of the page. Also, the lines start and end at positive and negative charges, respectively, unless they go off to infinity. Finally, the strength of the field in a given region is indicated by the density of lines there.



All of the basic principles here apply not only to collections of point charges, but also to continuous charge distributions. In order to find the net electric field due to such a distribution, we can chop it up into a bunch of tiny pieces which can be treated as point charges. Then by the principle of superposition,

$$\mathbf{E} = \int_V d\mathbf{E} = \int_V \frac{dq}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}},$$

where V is the volume inhabited by the charge distribution. Each dq is a function of the charge density at that point:

$$\text{in 1D} \quad dq = \lambda\, ds, \qquad \text{in 2D} \quad dq = \sigma\, dA, \qquad \text{and in 3D} \quad dq = \rho\, dV,$$

where λ, σ, ρ are the charge densities in their respective dimensions. Still, in general the integral is very difficult no matter the dimension, and we usually rely on a high degree of symmetry to solve problems. There are three main considerations here.

- Different distributions lend themselves particularly well to different coordinate systems. For an infinite line of charge we may choose to work in cylindrical coordinates, while for a spherically symmetric distribution we may instead use spherical coordinates.
- For infinite charge distributions, it's useful to see if symmetry allows for the cancellation of any field components. In the case of an infinitely long line of charge, we'd notice that $\mathbf{E}$ must only have a radial component and that, consequentially, the half-infinite lines of charge on either side of the field point have the same contribution to the net electric field.
- It's also useful to break up distributions in clever ways, in hopes of reducing the problem to a one-dimensional integral. For example, we might treat a charged sphere as a series of thin, concentric shells, each carrying a charge $\rho\, dV$, and integrate with respect to the distribution's radius.

Given how similar gravity is to the electrostatic force, it should be no surprise that Newton's shell theorem also applies to spherically symmetric charge distributions. All of the charge "interior" to the field point can be treated as one big point charge, while all of that "exterior" to the field point cancels and can thus be ignored.

1.2 Gauss's Law

Calculating the electric field via direct integration can be quite clunky. Fortunately, one of the fundamental equations of electromagnetism provides a much cleaner alternative. Gauss's law states that the electric flux through a closed surface S is proportional to the charge enclosed by the surface. Symbolically,

$$\oiint_S \mathbf{E} \cdot d\mathbf{A} = \frac{q_{\text{enc}}}{\epsilon_0},$$

where $d\mathbf{A}$ is an area element pointing along the outward normal and q_{enc} is the net charge enclosed by S. Although this statement technically holds true for any charge distribution, it's really only practical in a select few kinds of scenarios. The idea is to construct a fictitious "Gaussian surface" that respects the symmetry of the electric field: for planar symmetry we may use a Gaussian box, for cylindrical symmetry a Gaussian cylinder, and for spherical symmetry a Gaussian sphere.

Example: Electric field due to a plane of charge

Consider an infinite plane of charge with charge density σ. The corresponding electric field exhibits planar symmetry—all field lines point in the $\hat{z}$ (upward) direction, and the magnitude depends only on z—so upon an arbitrary region of the plane we construct a Gaussian box with some side length s.

To evaluate the surface integral, we take advantage of the fact that $\mathbf{E} = E\,\hat{z}$ is orthogonal to the "caps" of the box and parallel to the lateral faces:

$$\oiint_{\text{box}} \mathbf{E} \cdot d\mathbf{A} = \iint_{\text{caps}} \mathbf{E} \cdot d\mathbf{A} + \cancel{\iint_{\text{lats}} \mathbf{E} \cdot d\mathbf{A}} = 2 \iint_{\text{top}} E(|z|)\,\hat{z} \cdot dA\,\hat{z} = 2E(|z|) \iint_{\text{top}} dA,$$

so the integral evaluates to $E(|z|) \cdot 2s^2$. Since the box encloses a charge $q_{\text{enc}} = \sigma s^2$, by Gauss's law we have $E(|z|) \cdot 2s^2 = \sigma s^2$ and thus $\mathbf{E}(|z|) = (\sigma/2\epsilon_0)(\pm\hat{z})$, taking $+\hat{z}$ for positive z.

The integral form of Gauss's law states that electric flux is proportional to enclosed charge. This is useful, but there's another, more compact differential formulation. First note that, by the divergence theorem, we have

$$\iiint_V (\nabla \cdot \mathbf{E})\, dV = \oiint_{\partial V} \mathbf{E} \cdot d\mathbf{A}$$

for an arbitrary volume V and its boundary ∂V. By the integral form of Gauss's law the right-hand side turns into $q_{\text{enc}}/\epsilon_0$ or, alternatively,

$$\iiint_V (\nabla \cdot \mathbf{E})\, dV = \iiint_V \frac{\rho}{\epsilon_0} dV.$$

Since this equivalence is true for any volume we could think of, these integrands must be equal and we get

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0},$$

the differential form of Gauss's law. The "flux density" at a point is proportional to the charge density there.

1.3 Electrostatic Potential

Having discussed forces and fields, the next step we'll take in building a theory of electricity is to see how energy fits in. To exploit our understanding of the electrostatic force, we'll begin by computing the amount of work $\mathbf{F}_E$ does in moving a charge through an electric field. In particular, if q_0 is moved between points a and b while in the presence of another charge q, then in a spherical coordinate system centered at q we have

$$W_E = \int_a^b \mathbf{F}_E \cdot d\mathbf{s} = \int_a^b \frac{q_0 q}{4\pi\varepsilon_0 r^2} \hat{r} \cdot (dr\, \hat{r} + r\, d\theta\, \hat{\theta} + r \sin\theta\, d\phi\, \hat{\phi}) = -\left(\frac{q_0 q}{4\pi\varepsilon_0 b} - \frac{q_0 q}{4\pi\varepsilon_0 a} \right).$$

The electrostatic force is, like gravity, a conservative force and thus has an associated potential energy related to its work by $W_E = -\Delta U_E$. If we define $U_E(\infty) = 0$ like we often do for gravity we get

$$U_E(r) = \frac{q_0 q}{4\pi\epsilon_0 r}.$$

We might interpret this as the amount of work that must be done against $\mathbf{F}_E$ in order to bring q and q_0 into close proximity from infinite separation. Generalizing a bit, the total potential energy in a larger system of charges is the work required to construct the system by bringing each charge in from infinite separation, one by one. In a sense, the work is "stored" in the interactions between pairs of charges, and these interaction energies may be summed to determine the system's total electrostatic potential energy.

What we've essentially done here is turn a vector problem into a scalar one. Since $\mathbf{F}_E = -\nabla U_E$, in order to study the behavior of a particle under the influence of a net electrostatic force, we can simply look at the corresponding potential energy function and draw conclusions from there. This is often much more convenient than having to deal with vectors, and in fact we can do something similar with electric fields.

If an electric field represents an electric force per unit charge, then we can define a new quantity V, simply called potential (measured in volts V), to represent an electric potential energy per unit charge. Again assuming an infinite reference point, this means

$$V(r) = \frac{U_E(r)}{q_0} = \frac{q}{4\pi\epsilon_0 r},$$

for a point charge. Like all of the other quantities we've discussed so far, electrostatic potential obeys the principle of superposition. To find the potential due to a continuous distribution of charge we might again break the distribution of little chunks, approximate them as point charges, and integrate their individual contributions to the potential.

This also provides a new way to visualize electric fields. In two dimensions we might draw equipotential lines, along which the electrostatic potential is constant. The electric field, the negative gradient of this potential field, points normal to these lines in the direction of steepest descent.

Now, the difference in potential between two points a and b is unsurprisingly given by

$$-\Delta V = \int_a^b \mathbf{E} \cdot d\mathbf{s},$$

meaning we have the relationship $\mathbf{E} = -\nabla V$. Thus $\mathbf{E}$ is a conservative vector field and

$$\oint_C \mathbf{E} \cdot d\mathbf{s} = 0$$

for any closed C. This is a simplified version of Faraday's law (which we'll discuss in more detail later), and in differential form we can write $\nabla \times \mathbf{E} = \mathbf{0}$.

As a mathematical side note, we can also write $-\nabla \cdot (\nabla V) = \rho/\epsilon_0$. This is often expressed as

$$\nabla^2 V = -\rho/\epsilon_0$$

where ∇^2 is called the Laplacian operator, simply the sum of the second partial derivatives of V. We might loosely interpret this as the "net concavity" of V at a point—$\nabla^2 V$ is positive at points that are generally concave up, and negative at ones that are generally concave down. Since in free space we have $\nabla^2 V = 0$, V cannot have any extrema there. It also happens that, under any physically allowed boundary conditions, solutions to this equation are unique.

1.4 Conductors and Dielectrics

Now we'll take a look at how electric fields interact with two different kinds of matter: conductors and dielectrics. We'll begin with the simpler case of a conductor, which is a material in which each atom has at least one electron that's free to roam about. Arguably the most important feature of such a material is that its interior has no net electric field—if a conductor is placed into an external field $\mathbf{E}_0$, then its positive and negative charges will move around and create their own field $\mathbf{E}_{\text{ind}}$ in the opposite direction of $\mathbf{E}_0$. Eventually there will be an equilibrium in which $\mathbf{E}_{\text{ind}} = \mathbf{E}_0$.

There's a few other characteristics of conductors which often come in handy.

- If there is a cavity inside of a conductor, then it has no electric field. (Other there would be a C such that $\oint_C \mathbf{E} \cdot d\mathbf{s} \neq 0$.) Charge distributes itself over the cavity's surface to make this happen.
- Since there is no electric field in the interior of a conductor, by Gauss's law there is no net charge anywhere inside. Everything is distributed over the surface.
- The lack of electric field also means that all points inside of a conductor are at the same potential. To "ground" a conductor (i.e., set $V = 0$) we could connect it to a huge, faraway object, like the interior of the Earth.
- The electric field just outside of the conductor is orthogonal to the surface. Any tangential component would redistribute charge so as to cancel it out.

In other materials like dielectrics, electrons are generally bound to their molecules. However, under the influence of an external electric field, positive and negative charges can still be slightly displaced from one another. The precise mechanism by which this occurs depends on the kind of dielectric we're talking about: polar dielectrics have a bunch of permanent dipole moments that become aligned when influenced by an electric field, while nonpolar dielectrics have no such moments until being influenced. (In the latter case, electrons are simply displaced slightly from their normal position.)

Just like with conductors, the above rearrangement of charge induces an electric field that opposes the external field. In the case of linear dielectrics, the magnitude of this induced field $\mathbf{E}_{\text{ind}}$ is proportional to that of the applied field $\mathbf{E}_0$; for such materials we can define a dielectric constant

$$\kappa_E = \frac{E_0}{E} = \frac{E_0}{E_0 - E_{\text{ind}}},$$

which is roughly a measure of how conductive a material is. A perfect conductor would have $\kappa_E = \infty$. We can also see that electric fields inside of dielectrics are diminished by a factor of $1/\kappa$, which lines up with our intuition about conducting materials.

1.5 Capacitors

To cap off our discussion of electrostatics, we'll introduce a quantity that is fundamental to a variety of practical applications: capacitance. Suppose we have an isolated conductor with charge Q and a potential

V_0; then the capacitance C of the conductor, measured in farads (F), is the proportionality constant relating these quantities via $Q = CV_0$. More relevant to our purposes, however, is a definition for systems of two conductors with opposite charges $\pm Q$. Here we have $Q = C\Delta V$, where ΔV is the potential difference between the conductors. In either case, C represents the amount of charge that a system can carry per volt.

We'll focus on a special kind of capacitor comprised of two parallel plates, separated by a distance D, with equal and opposite charges $\pm Q$ distributed over an area A. For sufficiently large plates, we can exploit our Gauss's law result from earlier—we find that the plates' individual electric fields completely cancel outside of the capacitor and add to

$$E_{||} = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A}$$

on the inside. This turns out to be a great approximation for points that are not near the capacitor's fringe. Now, to find the capacitance of this arrangement we must first compute the potential difference between the two plates:

$$\Delta V = \left| \int_{\text{pos.}}^{\text{neg.}} \mathbf{E}_{||} \cdot d\mathbf{s} \right| = \left| \int_{\text{pos.}}^{\text{neg.}} E_{||}\, ds \right| = \frac{Q}{\epsilon_0 A} D.$$

This conveniently gives us a linear relationship between Q and ΔV, so we immediately get

$$Q = \left(\frac{\epsilon_0 A}{D} \right) \Delta V \implies C_{||} = \frac{\epsilon_0 A}{D},$$

a result which should make geometric sense in the limits for A and D.

Finally, we'll take a look at how much energy is stored in this parallel-plate capacitor. We could compute this directly using $\Delta U = Q\Delta V$, but we'll take a more enlightening approach. Suppose the plates are initially neutral; we can charge the capacitor by moving charges between the plates, one at a time, in total doing some amount of work

$$W = \int_0^Q (\Delta v)\, dq = \int_0^Q \left(\frac{q}{C} \right) dq = \frac{Q^2}{2C}.$$

Since the electric field is conservative we have $W = \Delta U$, meaning this is the total amount of energy stored in the capacitor! Some useful equivalent expressions are

$$U = \frac{Q^2}{2C} = \frac{C(\Delta V)^2}{2} = \frac{Q\Delta V}{2}.$$

We can go a step further. If we substitute $C = \epsilon_0 A/D$ and $\Delta V = E_{||} D$ into the latter expression, we get $U = (\epsilon_0 E_{||}^2 Ad)/2$. But Ad is simply the volume in which the electric field exists between the plates! So we can define a new quantity

$$u_E = \frac{U}{\text{vol.}} = \frac{1}{2}\epsilon_0 E^2,$$

called the "energy density" of an electric field. This suggests that we might think of the capacitor's energy as actually being stored in the field itself!

So we now have two ways of qualifying the energy stored in an electric field:

- the configuration energy, which describes the work done in building the electric field, and
- the energy density, which describes the amount of energy stored in the field per unit volume.

This also provides us with two ways to compute the energy stored in the electric field due to some charge distribution. We may assemble the configuration charge-by-charge, or we may simply integrate the energy density over all space.

2 Magnetostatics

2.1 Current

In our study of electrostatics, we assumed that all charges were basically stationary. In this static equilibrium, we have $E = 0$ and a lack of net charge inside of conductors. But in dynamic situations, charge may move more persistently!

Consider a wire with some charge flowing through it. The rate at which charge passes through a given cross section of the wire is called the current

$$i = \frac{dq}{dt}$$

at that point, measured in amperes (A). To account for scenarios in which the flow of charge is not uniform over the full cross section, we can define a current density $\mathbf{j}$ via

$$i = \iint \mathbf{j} \cdot d\mathbf{A},$$

where $\mathbf{j}$ points in the direction of positive charge flow. Of course we now know that it's actually negative charges (electrons) that're moving, but we stick to this convention because it isn't really hurting anyone.

Speaking of which, when we say that negative charge moves throughout a wire, we don't actually mean that electrons are rocketing through the wire at relativistic speeds, all in the same direction. In reality their motion is much more gas-like, bouncing around with high velocities that mostly cancel each other out. The component that does not cancel is called the drift velocity $\mathbf{v}_d$, and it tends to be quite slow, on the order of one millimeter per second. This is what creates a net movement of charge.

Now, in order to maintain this steady current throughout the wire, the electrons must be generally drifting at a constant speed. There are two things making this happen: there is a nonzero electric field in the wire pushing the electrons along, and there is some "drag" force keeping the electrons from accelerating. The forces due to these two cancel at the terminal velocity $\mathbf{v}_d$.

The drift velocity is related to the current density via the equation $\mathbf{j} = \rho \mathbf{v}_d$, where ρ is the charge density in the wire. If we make the simplest possible assumption that the drift velocity is proportional to the electric field, we get

$$\mathbf{j} = \sigma \mathbf{E}.$$

This is called Ohm's law. Specifically, this is its intensive form since it does not depend on any of the wire's geometric properties like length or radius. The quantity σ is called conductivity, and it quantifies how well a material can conduct current per unit electric field. (Its inverse σ^{-1} is called resistivity.)

This relationship is useful, but we can use it to derive something that might be more familiar. Suppose, now, that our wire has length L and cross-sectional area A. Since there is an electric field $\mathbf{E}$ in the wire, there is a potential difference ΔV between the two ends; assuming constant $\mathbf{E}$ and σ,

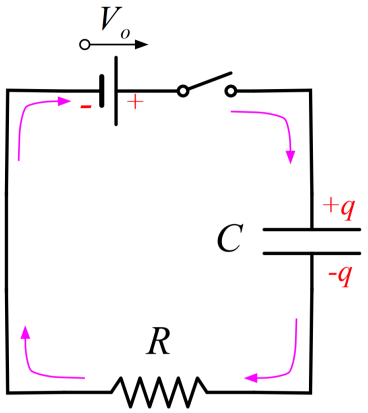
$$\Delta V = \left| \int_{x=0}^{x=L} \mathbf{E} \cdot d\mathbf{s} \right| = \left| \int_0^L \frac{j}{\sigma} dx \right| = \frac{jL}{\sigma}.$$

Since $j = i/A$, we have

$$\Delta V = i \left(\frac{L}{\sigma A} \right) = iR,$$

where we've defined the resistance R, measured in ohms (Ω). This is the extensive form of Ohm's law—note that R depends on the geometric properties of the wire. Devices that are created for the purpose of providing resistance are called resistors.

2.2 Circuits



Using wires, we can connect resistors, capacitors, and batteries (voltage sources) together to make current flow in ways that are practically useful to us. Schematics like the one at left are particularly useful abstractions when we want to work with circuits mathematically, like we will here.

Each circuit element is labeled with a corresponding quantity. The resistor has a resistance R, the capacitor a capacitance C, and the battery a voltage V_0. The battery dictates the direction of current flow—it pushes charge from low potential toward high potential, causing charge to move clockwise about the circuit. Right next to the battery is a switch; when closed current is allowed to flow, but when opened there is nowhere for the charge to go and the current ceases.

There are two key rules that we can exploit to "solve" a circuit.

- By conservation of energy, the net change in potential over any loop in the circuit must be zero. For these purposes we can ignore any drop in potential due to travel in a wire, since it will be negligibly small compared to the potential differences across the primary circuit elements.
- By conservation of charge, the total current flowing into any node (any place where two wires meet) must equal the current leaving it.

Using these two rules, we are usually able to construct a system of equations that, when solved, gives the current and potential difference across each circuit element present. We can also use them to find some convenient rules about combining like circuit elements together. For example, we could show that pairs of resistors in parallel or in series have respective equivalent resistances

$$R_{\text{par}} = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{-1}, \qquad R_{\text{ser}} = R_1 + R_2.$$

We can be a bit more sophisticated, too, and examine the behavior of circuits over time. This is especially useful in seeing, for example, how the charge on a capacitor changes while charging or discharging.

Example: Charging an RC circuit

Consider the above circuit. By conservation of energy, when the switch is closed we have

$$V_0 = iR + \frac{q}{C}.$$

We can express this as a differential equation in $q(t)$, the charge on the capacitor:

$$\frac{dq}{dt} + \frac{1}{RC}q = \frac{V_0}{R}.$$

With the initial condition $q(0) = 0$, the solution is

$$q(t) = CV_0\left(1 - e^{-t/RC}\right).$$

As more charge gets deposited onto the positive plate of the capacitor, it becomes more difficult to push charge off of the other plate, around the circuit, and back onto the positive plate.

Finally, in order to keep the current flowing, the battery must do work on the charge to move them to higher potentials. This work is, of course, given by $W = qV_0$, corresponding to a power

$$P_{\text{battery}} = \frac{dW}{dt} = iV_0.$$

The resistor does an equal amount of work, so by Ohm's law we have

$$P_{\text{resistor}} = \frac{V_0^2}{R} = i^2R$$

dissipated as heat.

2.3 Magnetic Fields

Experimental evidence has shown that electric fields do not exist in isolation. There is another field, called the magnetic field $\mathbf{B}$ (measured in teslas T), which is generated by moving charge. If we were to "grab" the wire with our right hand, our thumb sticking in the direction of current, then our fingers would curl in the direction of the magnetic field. We might verify this by looking at how the wire deflects a compass needle.

Not only are magnetic fields generated by moving charge, they also only act upon moving charge. In particular, if a charge q moves with velocity $\mathbf{v}$ within a magnetic field $\mathbf{B}$, then the magnetic force on the charge is

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}.$$

There are a few things to note here. Most importantly, since $\mathbf{F}_B$ is always orthogonal to $\mathbf{v}$, the magnetic force can never do any work! This perpendicularity also allows $\mathbf{F}_B$ to act as a centripetal force when $\mathbf{v} \perp \mathbf{B}$, creating circular or helical trajectories. Finally, if there are both electric and magnetic fields present, we add their effects to get the Lorentz force $\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}$.

Npw, since magnetic fields influence moving charge, they must also influence current-carrying wires. Suppose we have a segment of wire with current density $\mathbf{j}$ and electron drift velocity $\mathbf{v}_d$. If a constant magnetic field $\mathbf{B}$ acts on the wire, then the force on each electron is

$$\mathbf{F}_{B,e} = -q_e\mathbf{v} \times \mathbf{B} = -q_e\left(\frac{\mathbf{j}}{-q_e n}\right) \times \mathbf{B} = \frac{\mathbf{j} \times \mathbf{B}}{n},$$

where q_e is the elementary charge and n is the number density of electrons. If this segment of wire has N charges, then

$$\mathbf{F}_{B,L} = N\mathbf{F}_{B,e} = nAL\left(\frac{\mathbf{j} \times \mathbf{B}}{n}\right) = i\mathbf{L} \times \mathbf{B},$$

where L is the length of the segment, A is its cross-sectional area, and $\mathbf{L}$ points in the direction of current with magnitude L.

Now, everything we've done so far assumes that we already have $\mathbf{B}$ at all points in space. If we are instead only given a charge distribution, then we can find the magnetic field using the Biot-Savart law:

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{i d\mathbf{l} \times \hat{r}}{r^2},$$

where i is the current in a small segment $d\mathbf{l}$ of wire and $\mathbf{r}$ points from the wire segment to the field point. μ_0 is called the permeability of free space. As we'll see now, applying the Biot-Savart law is very similar to calculating electric fields via direct integration.

Example: Magnetic field due to a current-carrying ring

Consider a ring of wire with radius R and current i, which flows counterclockwise when viewed from above. We'll determine the magnetic field at a distance z above the center of the ring.

By the right-hand rule, chunks of wire that are opposite each other on the ring create a net magnetic field that points only in the vertical direction. This means we can focus solely on the magnetic field's $\hat{z}$ component:

$$dB_z = |d\mathbf{B}| \cos\theta = |d\mathbf{B}|\frac{R}{\sqrt{R^2 + z^2}},$$

where θ is the angle of inclination when z is viewed from a point on the ring. (This is equivalent to the angle each $d\mathbf{B}$ makes with the vertical.) By the Biot-Savart law we have

$$|d\mathbf{B}| = \frac{\mu_0}{4\pi}\left|\frac{i d\mathbf{l} \times \hat{r}}{r^2}\right| = \frac{\mu_0 i}{4\pi r^2}|d\mathbf{l} \times \hat{r}| = \frac{\mu_0 i}{4\pi r^2} dl.$$

Integrating dB_z to determine the total field:

$$\mathbf{B}_{\text{ring}} = \int_{\text{ring}} \frac{\mu_0 i\, dl}{4\pi(R^2 + z^2)}\frac{R}{(R^2 + z^2)^{1/2}}\hat{z} = \frac{\mu_0 i R}{4\pi(R^2 + z^2)^{3/2}}\hat{z}\int_{\text{ring}} dl = \frac{\mu_0 i \pi R^2}{2\pi(R^2 + z^2)^{3/2}}\hat{z}.$$

This current-carrying ring will actually become our model for a magnetic dipole. Just as we had the electric dipole moment qd, we have the magnetic dipole moment iA where A is the area enclosed by the ring. We can see that the magnetic field due to such a dipole obeys $B \propto iA/r^3$ for $r \gg R$, just like we saw with electric dipoles (in which case it was $E \propto qd/r^3$).

2.4 Ampere's Law

Just like with electric fields, this direct integration can get a little unwieldy. There's a cleaner way to determine magnetic fields in high-symmetry scenarios, but it'll be a bit different from Gauss's law. The main thing is that we have never detected any magnetic monopoles. This isn't a fundamental law of the universe or anything, it's just that every magnet we've encountered has had both a north pole and a south pole. Thus

$$\oiint_S \mathbf{B} \cdot d\mathbf{A} = 0, \qquad \nabla \cdot \mathbf{B} = 0.$$

Not very helpful for our purposes. But note that static magnetic fields are known for their twisty behavior about moving charges rather than the electric field's outward spray around point charges. So we can instead write down Ampere's law,

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \mu_0 i_{\text{enc}},$$

where i_{enc} is the net current puncturing an area enclosed by C. Once again, this statement holds for all scenarios involving magnetic fields, but it's only useful in a select few with the proper symmetries.

Example: Magnetic field due to a solenoid

Consider an infinite coil (or solenoid) of wire with radius R, current i, and n loops per unit length. We'll determine the magnetic field at some distance r from the central axis of the solenoid.

Let $\hat{z}$ represent the direction in which the magnetic field points, in accordance with the Biot-Savart law. The solenoid exhibits symmetry with respect to the cylindrical coordinates ϕ and z, so the magnetic field is only a function of r. Thus $\mathbf{B} = B(r)\,\hat{z}$.

Choose, as our Amperian loop, a length-r rectangle whose height h coincides with the solenoid's central axis. The loop is oriented such that it is parallel to $\hat{z}$ on the inside of the solenoid and antiparallel to $\hat{z}$ on the outside.

There are two regions to consider. For $r < R$, we have $i_{\text{enc}} = 0$. Also, since the lengths of the Amperian loop are orthogonal to the magnetic field, only the heights are relevant:

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = \int_{\text{up}} B(0)\,\hat{z} \cdot dl\,\hat{z} + \int_{\text{down}} B(r)\,\hat{z} \cdot dl(-\hat{z})$$

Now, it can be shown by a direct integration of $\mathbf{B}_{\text{ring}}$ that the magnetic field at the center of an infinite solenoid is $\mathbf{B}(0) = \mu_0 ni$. Thus by Ampere's law

$$0 = \mu_0 nih - B(r)h \implies \mathbf{B}_{\text{in}}(r) = \mu_0 ni\,\hat{z}.$$

So the magnetic field inside of a solenoid is constant. For $r > R$ the setup is the same, only now we have $i_{\text{enc}} = nhi$. (i_{enc} is positive here due to how we've oriented our Amperian loop.) Thus by Ampere's law,

$$\mu_0 nih - B(r)h = \mu_0 nhi \implies \mathbf{B}_{\text{out}}(r) = \mathbf{0}.$$

2.5 Frames of Reference

At this point we have developed a complete theory of statics. Everything we've done so far can be derived from Maxwell's time-independent equations in their integral and differential forms:

$$\oiint \mathbf{E} \cdot d\mathbf{A} = \frac{q}{\epsilon_0} \qquad \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \qquad \nabla \times \mathbf{E} = \mathbf{0}$$

$$\oiint \mathbf{B} \cdot d\mathbf{A} = 0 \qquad \nabla \cdot \mathbf{B} = 0$$

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{j}$$

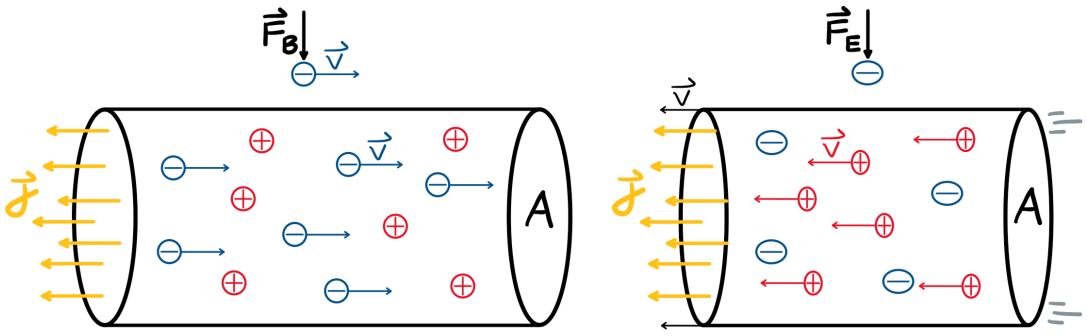
There is more to the theory of electromagnetism—we'll soon examine what happens when $\mathbf{E}$ and $\mathbf{B}$ are allowed to vary with time—but first we'll take a closer look at how these two fields are related, and why we so often see them lumped together as a unified electromagnetic field.

Consider the scenario illustrated below, on the left. In a neutral wire with cross-sectional area A electrons are free to move with drift velocity $\mathbf{v}$, generating a current density $\mathbf{j}$ in the opposite direction. We could use Ampere's law to show that the magnetic field due to this wire is $B = \mu_0 jA/2\pi r$, with direction determined by the right-hand rule.

Outside of the wire there is an electron that also has velocity $\mathbf{v}$. The force on this electron is purely magnetic in nature; specifically, taking the standard x-y coordinate system,

$$\mathbf{F}_B = -q_e \mathbf{v} \times \mathbf{B} = q_e v \left(\frac{\mu_0 jA}{2\pi r} \right) (-\hat{\mathbf{y}}) = \frac{q_e v^2 \mu_0 \rho_- A}{2\pi r} (-\hat{\mathbf{y}}),$$

where we've defined the density of negative charge ρ_-, a positive quantity, for later convenience. Let us call this scenario the S frame.



Now suppose we, the observer, started walking with velocity $\mathbf{v}$. We now observe the electrons to be (in net) at rest, and it is the positive charges that are actually moving. Reason dictates that the isolated electron should still experience the same downward force, but this force cannot be magnetic since it's no longer in motion. We must therefore uncover some sort of hidden electric force in this new S' frame.

Special relativity, specifically length contraction, provides an answer! When compared to the S frame, in the S' frame the positive charges are contracted while the electrons are uncontracted. So while in the unprimed frame we had $\rho_{\text{net}} = \rho_+ - \rho_- = 0$, in the primed frame we see

$$\rho'_{\text{net}} = \rho'_+ - \rho'_- = \gamma \rho_+ - \gamma^{-1} \rho_- = \gamma \frac{v^2}{c^2} \rho_-,$$

where γ is the Lorentz factor $\left(1 - v^2/c^2\right)^{-1/2}$. So the lone electron will at least experience an attractive force, which is promising. To find its magnitude we appeal to Gauss's law: a Gaussian cylinder with radius r' and length l' encloses a charge $\rho' A' l' / \epsilon_0$ which generates a flux $E'(2\pi r l')$, meaning

$$E' = \frac{\rho' A'}{2\pi \epsilon_0 r'} = \gamma \frac{v^2 \rho_- A}{2\pi \epsilon_0 c^2 r} \implies \mathbf{F}'_E = \gamma \frac{q_e v^2 \rho_- A}{2\pi \epsilon_0 c^2 r} (-\hat{\mathbf{y}}).$$

Now we can compare the two forces. We could use time dilation to argue that $\mathbf{F}_E = \gamma^{-1}\mathbf{F}'_E$, so we have

$$\mathbf{F}_B = \mu_0 \frac{q_e v^2 \rho_- A}{2\pi r}(-\hat{\mathbf{y}}), \qquad \mathbf{F}_E = \frac{1}{\epsilon_0 c^2}\frac{q_e v^2 \rho_- A}{2\pi r}(-\hat{\mathbf{y}}).$$

We can see that these expressions are equivalent if $\epsilon_0\mu_0 = 1/c^2$ and, remarkably, experimental evidence suggests that this is actually the case! Thus a force that appeared as purely magnetic in the S frame is actually purely electric in the S' frame.

So it doesn't really make sense to speak about electric and magnetic fields as two separate entities. They are, in fact, two facets of a more fundamental electromagnetic field, related to each other via a Lorentz transformation:

$$\begin{aligned}
\mathbf{E}'_{||} &= \mathbf{E}_{||} \\
\mathbf{E}'_{\perp} &= \gamma\left(\mathbf{E}_{\perp} + \mathbf{v}\times\mathbf{B}\right) \\
\mathbf{B}'_{||} &= \mathbf{B}_{||} \\
\mathbf{B}'_{\perp} &= \gamma\left(\mathbf{B}_{\perp} - \frac{1}{c^2}\mathbf{v}\times\mathbf{E}\right)
\end{aligned}$$

Visually, at relativistic frame velocities the electric and magnetic fields get compressed along the direction of motion, similar to how lengths are contracted at these speeds.

3 Electrodynamics

3.1 Faraday's Law

Now it's time to bring time-dependence into the picture by allowing electric and magnetic fields to change over time. Before we begin, though, we should take a closer look at the actual phenomenon responsible for current in wires.

There are two things that contribute to driving current around a loop of wire: a source $\mathbf{f}$ that does the pushing and an electrostatic field that smooths the current flow throughout the wire. The source can take a variety of forms—we often think of it as a chemical force confined to the inside of a battery—but whatever it is, its net effect is characterized by the electromotive force (or emf),

$$\mathcal{E} = \oint \mathbf{f} \cdot d\mathbf{l}.$$

Note that, confusingly, $\mathcal{E}$ is not a force but rather the integral of a force per unit charge. This means emf has units of potential, and we can often interpret it as such, like in the case of an ideal battery. The following example illustrates a case in which this is not so.

Example: Motional emf

Consider a height-h rectangular loop of wire whose left half is enveloped by a uniform magnetic field $\mathbf{B}$ pointing into the page. Suppose we pull the loop toward the right with velocity $\mathbf{v}$; an upward magnetic force acts on the charges in the "vertical wire segment", generating a clockwise current.

The magnetic force has done no work here (the extra energy comes from our pull), it still creates an emf

$$\mathcal{E} = \oint_{\text{wire}} (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} = vBh.$$

This simple example is indicative of a more general rule. Whenever the magnetic flux Φ through a loop of wire changes, we have

$$\mathcal{E} = -\frac{d\Phi}{dt} = -\frac{d}{dt} \iint \mathbf{B} \cdot d\mathbf{A}.$$

It turns out that this is our new, time-dependent version of Faraday's law. We'll come back to this in a moment. Now, although the above equation leaves no ambiguity regarding the direction of the induced emf (it is determined by the right-hand rule), the signs can get confusing. To get the directions right, we introduce Lenz's law: the induced current will flow in such a way that opposes the change in magnetic flux. If, for example, a magnetic field into the page is increasing in magnitude, then current will flow counterclockwise to create a magnetic field out of the page.

The above form of Faraday's law makes sense if the emf is induced by the wire's motion through space, but experiment shows that it also holds for stationary wires in changing magnetic fields. In this case the emf cannot be magnetic in nature, so it must be electric. Specifically, a changing magnetic field $\mathbf{B}$ induces an electric field $\mathbf{E}$ that satisfies

$$\oint_{\partial S} \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \iint_S \mathbf{B} \cdot d\mathbf{A}.$$

This is the modified version of Faraday's law that we'll include in Maxwell's equations. We could use Stokes's theorem to get the differential form,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.$$

3.2 Inductors

One surprising consequence of Faraday's law is that it allows a circuit to induce an emf in itself. Consider an open, clockwise-oriented circuit that is isolated from any external magnetic fields; when the circuit is closed, current flows around the wire and generates a magnetic flux into the page. But this generation isn't instantaneous, meaning there is a period where $d\Phi/dt$ is nonzero and, consequentially, where there is an emf opposing the change in current.

This phenomenon is called self-inductance, and the opposing emf is called back-emf. The magnitude of the back-emf is determined by the inductance L of the circuit, defined by

$$\mathcal{E} = -L\frac{di}{dt}.$$

Inductance is measured in henries (H) and depends on the relevant circuit's material and geometric properties.

Example: Inductance of a solenoid

Consider a radius-r solenoid with n turns per unit length, spanning a length l. A current i flows through the solenoid. The resulting magnetic flux is given by

$$\Phi = \iint_{\text{sol}} \mathbf{B}\cdot d\mathbf{A} = nl \iint_{\text{loop}} \mathbf{B}\cdot d\mathbf{A} = nl\cdot \mu_0 ni\cdot \pi r^2.$$

By Faraday's law, the induced emf produced by a changing current has magnitude

$$|\mathcal{E}| = \left|\frac{d}{dt}\left(nl\cdot\mu_0 ni\cdot\pi r^2\right)\right| = \mu_0 n^2(\pi r_s^2 l)\cdot\left|\frac{di}{dt}\right|.$$

Thus the inductance of the solenoid is

$$L_{\text{sol}} = \mu_0 n^2\left(\pi r_s^2 l\right),$$

where A is the area of each loop.

Solenoid-like objects often make appearances in electrical circuits as inductors—they serve to increase the effective area enclosed by the circuit, generating a stronger back-emf in responses to changes in current. Since the area created by an inductor is usually much greater than the other area enclosed by the circuit, for our purposes we can model the entire circuit's inductance with a single inductor.

Like every other circuit element, inductors have potential differences across their terminals. It is simply given by $\mathcal{E}$—when the current is increasing there is a potential drop, and when current is decreasing there is a rise. Like we have in the past, we can use this with conservation of energy to construct a differential equation that describes the current dynamics in a circuit.

Inductors store energy in the magnetic field in a fashion analogous to how capacitors store energy in the electric field. The flow of energy into an inductor is given by

$$P = i\mathcal{E}_{\text{ind}} = iL\frac{di}{dt},$$

so the overall work done in charging the conductor is

$$W = \int P\,dt = \frac{1}{2}Li^2.$$

This is equivalent to the amount of "magnetic" potential energy U_B stored in the inductor. We can also define an energy density by determining the energy per unit volume stored in a solenoid:

$$u_B = \frac{\frac{1}{2}L_{\text{sol}}i^2}{\pi r_{\text{sol}}^2 l_{\text{sol}}} = \frac{1}{2\mu_0}B^2.$$

This gives us another way to determine the inductance of an object—given a magnetic field and a current, we can integrate an energy density over the object's volume to find its stored energy and solve for its inductance.

3.3 Maxwell's Equations

We've found that changing magnetic fields generate electric fields. But in order to preserve the relativistic symmetry between the two fields, we should also expect that changing electric fields generate magnetic fields. We'll use a thought experiment to resolve this discrepancy!

Suppose a current i charges a standard parallel-plate capacitor. We'll draw an Amperian circle around the current, accompanied by two different surfaces S_1 and S_2 with that circle as their mutual boundary.

- S_1 is simply the filled-in disk corresponding to the Amperian circle. It encloses a current i, so by Ampere's law we have $B = \mu_0 i/2\pi r$, where r is the radius of the Amperian circle.
- S_2 is an open cylinder whose cap is between the plates of the capacitor. Since this surface encloses no current, our version of Ampere's law gives $B = 0$.

These two cases are clearly at odds with each other—Ampere's law should give the same result no matter what surface we use. We might propose that the capacitor's time-varying electric flux can solve the problem. This would give an equation of the form

$$\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \mu_0 i_{\text{enc}} + ? \frac{d}{dt} \iint_S \mathbf{E} \cdot d\mathbf{A},$$

where $?$ represents a constant that we'll now solve for. Noting that the electric field between the capacitor plates is $E = q/(A\epsilon_0)$, where A is the area of each plate, using S_2 as our surface we get from the above

$$\begin{aligned} B(r) \cdot 2\pi r = ? \frac{d}{dt} \iint_{\text{cap}} \mathbf{E} \cdot d\mathbf{A} = ? \frac{d}{dt} \frac{q}{A\varepsilon_0} A = \frac{?}{\varepsilon_0} \frac{dq}{dt}. \\ B(r) = \frac{? \, i}{\varepsilon_0 \cdot 2\pi r} \end{aligned}$$

So $B(r) = (?i)/(2\pi r \epsilon_0)$, and in order to be consistent with our result for S_1 we must have $? = \mu_0 \epsilon_0$. We've resolved the discrepancy! From here we could again use Stokes's theorem to derive the differential form. Note that Ampere's law can be written as

$$\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \left(i_{\text{enc}} + \epsilon_0 \frac{d}{dt} \iint_S \mathbf{E} \cdot d\mathbf{A} \right),$$

so the latter term in the parentheses has units of current. Confusingly, this term is called the displacement current despite having nothing to do with actual current at all. Another example of vestigial jargon.

Anyway, we've finally arrived at the complete set of Maxwell's equations, the fundamental principles of classical electromagnetism.

$$\begin{aligned} \oiint_S \mathbf{E} \cdot d\mathbf{A} &= \frac{q_{\text{enc}}}{\varepsilon_0} & \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \oint_{\partial S} \mathbf{E} \cdot d\mathbf{l} &= -\frac{d}{dt} \iint_S \mathbf{B} \cdot d\mathbf{A} & \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \oiint_S \mathbf{B} \cdot d\mathbf{A} &= 0 & \nabla \cdot \mathbf{B} &= \mathbf{0} \\ \oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} &= \mu_0 i_{\text{enc}} + \mu_0 \varepsilon_0 \frac{d}{dt} \iint_S \mathbf{E} \cdot d\mathbf{A} & \nabla \times \mathbf{B} &= \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \end{aligned}$$

Everything we know about electricity and magnetism is embedded in these laws. For example, we can take the divergence of Ampere's law to get charge conservation:

$$\begin{aligned} \nabla \cdot \nabla \times \mathbf{B} &= \nabla \cdot \left(\mu_0 \mathbf{j} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \\ 0 &= \nabla \cdot \mu_0 \mathbf{j} + \nabla \cdot \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \\ 0 &= \nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} \end{aligned}$$

In words, the current density flowing into or out of a point is equal and opposite the change in charge density.

3.4 Electromagnetic Waves

Let's see what happens when we take the curl of Faraday's law. Applying a vector calculus identity gives

$$\nabla \times \nabla \times \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B})$$
$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{\partial}{\partial t}\left(\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right)$$
$$\nabla^2 \mathbf{E} = \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2}$$

We recognize this as the wave equation in $\mathbf{E}$! (A similar line of reasoning for magnetic fields gives the wave equation in $\mathbf{B}$.) Recall that solutions to this equation take the form $\mathbf{E}(x, y, z, t) = \mathbf{E}_0 f(x \pm ct)$, which corresponds to a wave propagating in the $\mp x$-direction with speed c. This is what we call light!

These electromagnetic waves come with some nice properties. If we have an electric field $\mathbf{E} = \mathbf{E}_0 f(x - ct)$, then

$$\nabla \cdot \mathbf{E} = E_{0x}\frac{\partial}{\partial x} f(x - ct) = E_{0x} f'(x - ct).$$

By Gauss's law this must evaluate to zero. It doesn't really make sense for a wave in motion to have $f'(x - ct) = 0$, so we must have E_{0x}. Thus $\mathbf{E}$ cannot have a component along the direction of wave propagation! A similar analysis would show that $\mathbf{B}$ cannot have such a component, either.

Let's take a closer look at the relationship between the two fields. If we take $\mathbf{E}$ to point in the $\hat{y}$ direction, then we'd get a curl

$$\nabla \times \mathbf{E} = \hat{z}\,\frac{\partial}{\partial z} E_0 f(x - ct);$$

by Faraday's law this is equivalent to $-\partial \mathbf{B}/\partial t$, meaning the electric and magnetic fields are orthogonal to one another! Specifically, if we have $\mathbf{B} = B_0 \hat{z}\, f(x - ct)$, then taking a time derivative yields

$$\frac{\partial \mathbf{B}}{\partial t} = -\hat{z}\, c B_0 f'(x - ct),$$

just as we got above. Since $\hat{y} \times \hat{z} = \hat{x}$, this means $\mathbf{E} \times \mathbf{B}$ points in the electromagnetic wave's direction of propagation. We also get a relationship between the amplitudes of the fields,

$$B_0 = \frac{E_0}{c}.$$

If we wanted, we could verify that all this theory has remained consistent with Maxwell's equations. To verify Faraday's integral law we might take a tall, thin loop in the plane of $\mathbf{E}$ and qualitatively argue that the sign of the line integral about the loop is opposite that of the corresponding change in magnetic flux. For Ampere's law we'd do something similar with a loop in the plane of $\mathbf{B}$.

3.5 The Poynting Vector

Aided by a general understanding of electromagnetic waves, we're ready to take a more detailed look at how energy fits into the broad picture of electrodynamics. We'll begin with the simple observation that the total energy stored by the electromagnetic field is given by

$$U_T = \iiint (u_E + u_B)\, d(\text{vol.}) = U_E + U_B = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2\right).$$

In the special case of an electromagnetic wave we have $B^2 = \mu_0 \epsilon_0 E^2$, which gives $U_T = \epsilon_0 E^2$. So in such a wave, the field energy is equally electric and magnetic!

Now we'll look at the dynamics of this total field energy. We'll first dot the electric field onto both sides of Ampere's law and apply a couple of vector identities:

$$\mathbf{E} \cdot (\nabla \times \mathbf{B}) = \mu_0 \varepsilon_0\, \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t},$$
$$\mathbf{B} \cdot (\nabla \times \mathbf{E}) - \nabla \cdot (\mathbf{E} \times \mathbf{B}) = \frac{\mu_0 \varepsilon_0}{2}\frac{\partial}{\partial t} E^2.$$

By Faraday's law,

$$-\nabla \cdot (\mathbf{E} \times \mathbf{B}) = \frac{1}{2}\mu_0\varepsilon_0 \frac{\partial}{\partial t} E^2 + \mathbf{B} \cdot \left(\frac{\partial}{\partial t} \mathbf{B} \right),$$
$$-\nabla \cdot \left(\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) = \frac{1}{2}\varepsilon_0 \frac{\partial}{\partial t} E^2 + \frac{1}{2\mu_0} \frac{\partial}{\partial t} B^2.$$

We might recognize this as

$$-\nabla \cdot \left(\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) = \frac{\partial}{\partial t} u_T,$$

which we can integrate to get

$$\oiint_{\partial V} \left(\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \right) \cdot (-d\mathbf{A}) = \frac{d}{dt} \iiint_V u_T \, d(\text{vol.}).$$

If we look closely, we might see that this is a statement about conservation of energy! The right side is the rate at which the "electromagnetic energy" inside a volume V changes, so we can infer that the left side describes the inward flux of energy through the boundary ∂V. Thus the integrand

$$\mathbf{S}_P = \frac{\mathbf{E} \times \mathbf{B}}{\mu_0},$$

called the Poynting vector, describes the energy transported by the electromagnetic field per unit time, per unit area. (We might call $\mathbf{S}$ an energy flux density.) This relationship is fundamental to the study of energy in electrodynamics, so naturally we can use it to do some pretty cool things! For example, we could use it to prove that the power emitted by a resistor is given by i^2R.

Electromagnetic waves, of course, carry energy, so the Poynting vector naturally coincides with an such a wave's direction of propagation. But while we usually think of a light beam as having a fixed energy, in reality the magnitude of $\mathbf{S}_P$ oscillates rapidly between extrema. So we define the intensity I of an electromagnetic wave as the magnitude of its time-averaged Poynting vector; since $\langle \sin^2(x - ct) \rangle = 1/2$, we have

$$I = \langle S_P \rangle = \left\langle \frac{EB}{\mu_0} \right\rangle = \left\langle \frac{E_0^2 \sin^2(x - ct)}{\mu_0 c} \right\rangle = \frac{1}{2\mu_0} \frac{E_0^2}{c} = \frac{1}{2}\epsilon_0 c E_0^2.$$

More surprisingly, we know from special relativity that light carries a momentum $p = E/c$. We can use this to define radiation pressure, the force F an electromagnetic wave exerts on a surface per unit area A. By Newton's second law,

$$\text{pressure} = \frac{1}{A}\frac{dp}{dt} = \frac{1}{cA}\frac{dE}{dt} = \begin{cases} S/c \text{ (absorber)} \\ 2S/c \text{ (reflector)} \end{cases}.$$

The separate cases here arise from conservation of momentum. Absorbers simply take on the momentum of any incident light, while reflectors must double that amount in order to turn the light around.

3.6 Magnetism in Materials

We'll cap off our study of electrodynamics by looking at how magnetic fields interact with matter. So far we've studied these fields in the context of macroscopic charge movements—through a wire, for example—but we know from real-world experience that magnets also arise when there is no obvious movement of charge. In cases like these, we must zoom in to the microscopic level.

Electrons are known to have spin angular momentum. For our purposes, we may model this by having each electron spin about a central axis, generating a little loop of current i and thus a dipole moment $\boldsymbol{\mu} = i\mathbf{A}$. Electrons that are "paired" with one another cancel each others' dipole moments, but unpaired electrons are left to act as permanent dipole moments.

When a collection of these permanent dipoles is isolated from external magnetic fields, they all point in random directions that tend to cancel each other out, so no magnetism arises. But under the influence of a

uniform field $\mathbf{B}$, there is a net torque that tries to orient each loop such that $\boldsymbol{\mu}$ coincides with the direction of $\mathbf{B}$. We could show (quite easily for rectangles) that the torque on each loop is given by $\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}$, which gives the configuration energy $U_B = -\boldsymbol{\mu} \cdot \mathbf{B}$ if the loop does not move. In practice the loop does move, and all of the dipoles become aligned with the external magnetic field in a phenomenon called paramagnetism.

Now, electrons are known to also have orbital angular momentum, which we'll model as literal orbits generating slightly larger loops of current. Paired electrons are modeled by two loops of wire with equal opposite currents, producing no net magnetic dipole in isolation. When influenced by an external magnetic field, by Lenz's law the loops of wire together create a magnetic field that opposes the induced field; in aggregate, this phenomenon is called diamagnetism.

Magnetic fields are reduced in diamagnetic materials. To quantify this reduction, we can define a "diamagnetic constant" κ_m in a similar fashion to the dielectric constant: if $\mathbf{B}_0$ is the applied field and $\mathbf{B}$ is the resultant field, then $\mathbf{B} = \kappa_m \mathbf{B}_0$. For diamagnetic materials we have $\kappa_m < 1$, while for paramagnetic materials $\kappa_m > 1$. (Note that all materials are at least somewhat diamagnetic, but its effect is often very weak.)

Finally, when we take a paramagnet and allow the permanent dipoles to interact with one another in such a way that they remain aligned after removing the external magnetic field, we get a ferromagnet. These materials have $\kappa_m \gg 1$. Since interactions between atoms tend to decrease with temperature, each ferromagnet has a Curie temperature above which they become paramagnetic.

4 Optics

4.1 Light and Materials

Now we'll turn our attention to a study of light in its own right. To begin, we'll need to modify Maxwell's equations a little bit. Recall that they were constructed in free space, so in order to make them valid in materials we must substitute $\epsilon_0 \to \kappa_E \epsilon_0$ and $\mu_0 \to \kappa_B \mu_0$. Consequentially, in matter we have the wave equation

$$\nabla^2 \mathbf{E} = (\kappa_e \kappa_m) \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

If we define the material's index of refraction $n = \sqrt{\kappa_e \kappa_m}$, then we find that light travels at

$$v = \frac{1}{\sqrt{\kappa_e \kappa_m}} \frac{1}{\sqrt{\mu_0 \epsilon_0}} = \frac{c}{n}$$

when in a medium. It follows that electromagnetic waves satisfy $B_0 = E_0 n/c$, and that their intensities are given by $I = E_0^2 n / 2\mu_0 c$.

A point at which materials with different refractive indices meet is called an interface. When an electromagnetic wave is incident on an interface, it gets split into a reflected wave and a transmitted wave. For normal incidence, the equations describing this split are

$$\begin{aligned}
\mathbf{E}_i &= E_i \, \hat{y} \, \cos(k_1 x - \omega_1 t), & \mathbf{B}_i &= \frac{E_i n_1}{c} \, \hat{z} \, \cos(k_1 x - \omega_1 t), \\
\mathbf{E}_r &= E_r \, \hat{y} \, \cos(-k_1 x - \omega_1 t), & \mathbf{B}_r &= \frac{E_r n_1}{c} \, (-\hat{z}) \, \cos(-k_1 x - \omega_1 t), \\
\mathbf{E}_t &= E_t \, \hat{y} \, \cos(k_2 x - \omega_2 t), & \mathbf{B}_t &= \frac{E_t n_2}{c} \, \hat{z} \, \cos(k_2 x - \omega_2 t),
\end{aligned}$$

where n_1 is the refractive index of the first material and n_2 is that of the second. We can relate all of these equations using Ampere's law and Faraday's law; if we take thin loops that are parallel to $\mathbf{E}$ and $\mathbf{B}$, respectively, we get

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0, \qquad \oint \mathbf{B} \cdot d\mathbf{l} = 0.$$

So if we have an interface at $x = 0$, then the relationship between the electric fields is

$$E_i \cos(-\omega_1 t) + E_r \cos(-\omega_1 t) - E_t \cos(-\omega_2 t) = 0.$$

The only way for this equation to be true for all t is if $\omega_1 = \omega_2 - \omega$. meaning the frequency of light does not change in a medium! All that changes, then, is the wavenumber $k = \omega/v = 2\pi n / \lambda_{\text{vac}}$.

Anyway, cancelling the cosines gives $E_i + E_r = E_t$. A similar analysis for magnetic fields would give $E_i n_1 - E_r n_1 = E_t n_2$, and solving the resulting system of equations gives

$$E_r = \left(\frac{n_1 - n_2}{n_1 + n_2} \right) E_i, \qquad E_t = \left(\frac{2 n_1}{n_1 + n_2} \right) E_i.$$

Note that E_r is negative when $n_1 < n_2$, so in this case the reflected wave is out of phase with the transmitted wave. Otherwise the waves are in phase. Here's a handy mnemonic to keep track:

"Low to high, add a π." "High to low, let it go."

For scenarios with multiple interfaces we simply apply these rules several times, often ignoring light that bounces back and forth between interfaces many times.

4.2 Polarization

The orientation of an electromagnetic wave, called its polarization, can be described entirely by the direction in which its electric field points. Most common sources of light produce unpolarized light, in which the direction of $\mathbf{E}$ is basically random and time-varying. When light is polarized, though, it can come in three different forms: linear, circular, and elliptical.

- For light that is linearly polarized, $\mathbf{E}$ points in a constant direction as the wave propagates. A wave propagating in the x-direction may be polarized along $\hat{y}$, $\hat{z}$, or some linear combination of the two.
- Circularly polarized light is comprised of two identical components that are out of phase by precisely a quarter-cycle, so that the electric field appears to trace out a circle in space. Confusingly, optics people tend to specify the "handedness" of this tracing with respect to the receiving end of the light, so we point our thumb opposite the direction of propagation and curl our fingers in such a way that matches the twist of the $\mathbf{E}$-field. This gives us the following parameterizations.

$$\text{RHCP: } \mathbf{E} = E_0\,\hat{y}\cos(kx-\omega t) + E_0\,\hat{z}\sin(kx-\omega t)$$
$$\text{LHCP: } \mathbf{E} = E_0\,\hat{y}\sin(kx-\omega t) - E_0\,\hat{z}\cos(kx-\omega t)$$

- Elliptical polarization occurs we have two identical components that are out of phase by a different amount, so that the electric field appears to trace out an ellipse in space. Linear and circular polarization can be seen as special cases of elliptical polarization.

The polarization of light can be controlled using filters called polarizers. In the case of a linear polarizer, the incident electric field is split into two components $\mathbf{E}_{\text{in}} = \mathbf{E}_{\parallel} + \mathbf{E}_{\perp}$ which are parallel and perpendicular to the direction of polarization, respectively. As $\mathbf{E}_{\text{in}}$ passes through, $\mathbf{E}_{\parallel}$ is transmitted entirely while $\mathbf{E}_{\perp}$ is absorbed entirely. So if $\mathbf{E}_{\text{in}}$ is polarized at an angle θ with respect to the polarization direction, then $E_{\text{out}} = E_{\text{in}}\cos\theta$ and we have Malus's law

$$I_{\text{out}} = I_{\text{in}}\cos^2\theta.$$

Linear polarizers are the basis for a host of technologies, like polarizing sunglasses and LCD screens! They also exhibit some strange behavior due to Malus's law. Two polarizers oriented perpendicular to one another will block out all light, but if a third is added at a 45° angle to both of them, then suddenly some light will pass through.

There are other ways light can be polarized, too. One is through scattering: if some light traveling in the $\hat{z}$-direction and is incident on a collection of dust, then this dust is excited and begins to vibrate in the xy-plane (i.e., in the plane of the electromagnetic field). When light is emitted its propagation direction and polarization direction both remain in this plane and, of course, are orthogonal to each other.

Polarization by reflection may also occur when light reflects on an interface between materials. Define the light's plane of incidence to be that spanned by the vector normal to the interface and that which points in the direction of incident light. For any interface, the incidence-plane component of the reflected electromagnetic wave is dependent not only on the materials' refraction indices, but also on the angle of incidence. Thus there is a critical angle, called the Brewster's angle, such that all light polarized in the plane of incidence is transmitted. (This also happens to be the angle at which the reflected and transmitted light are orthogonal to one other.) As a result, the reflected light is polarized orthogonal to the plane of incidence.

Finally, polarization can occur when light passes through a birefringent material, the details of which don't concern us. In the big picture, some materials have refractive indices that depend on the direction in which light propagates through them. Such a material has an intrinsic optic axis with index n_o, while the perpendicular axis has n_e; the perpendicular components of the normally incident wave (the o- and e-rays) undergo a relative phase shift

$$\Delta\phi = \frac{2\pi}{\lambda}(n_o - n_e)d,$$

where λ is the light's wavelength and d is the width of the material. If $\Delta\phi = \pi/4$ or a coterminal equivalent, then we have a quarter-wave plate that can be used to convert 45°-incident linearly polarized light into circularly polarized light, or vice versa. In the case of $\Delta = \pi/2$ we have a half-wave plate which "mirrors" the polarization of the incident light.

4.3 The Photon

In classical optics, light is comprised of electromagnetic waves and has intensity given by the electric field, vacuum permeability, and speed of light. But we now know light to exhibit a kind of wave-particle duality—in quantum optics it is comprised of photons, each of which has an energy $E = h\nu$, where ν is the frequency of the light and h is called Planck's constant. So if a beam of N photons is incident on an area A in a time Δt, then the intensity of the light beam they comprise is given by

$$I = \frac{Nh\nu}{A\Delta t}.$$

The quantum nature of light has been experimentally verified several times using various experiments. One examined the photoelectric effect, the observation that electromagnetic radiation can "kick" electrons off the surface of a metal. These electrons leave the metal with, at most, a kinetic energy $K_{\max} = h\nu - W$, where W is the energy expended in escaping the metal to begin with, called the metal's work function. In the historical experiment, the metal is the high-potential terminal of an uncharged parallel-plate capacitor held in a vacuum chamber. This capacitor is part of a larger circuit so that, if the ejected electrons have a high enough kinetic energy, they can make it to the other side of the capacitor and drive a current. The critical potential where no electrons make it to the other terminal is called the stopping potential V_{stop}, and we have

$$q_e V = h\nu - W.$$

Experiments have verified this, meaning the stopping potential has absolutely nothing to do with the intensity of the incident light. This opposes the classical theory that would have V_{stop} increase with intensity.

In another experiment, high-energy light is scattered off of a graphite target. Assuming light is quantized, we'll see a photon with wavelength λ collide with an electron of mass m_e; the photon is deflected by an angle θ with a new wavelength λ', while the electron is sent off at an angle ϕ with momentum p_e. Using relativistic conservation of momentum and energy we get the equations

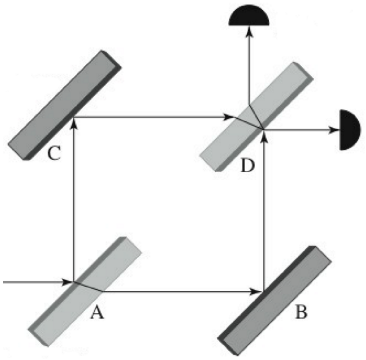
$$\frac{h}{\lambda} = \frac{h}{\lambda'}\cos\theta + p_e\cos\phi, \qquad 0 = \frac{h}{\lambda'}\sin\theta - p_e\sin\phi,$$
$$m_e c^2 + \frac{hc}{\lambda} = \frac{hc}{\lambda'} + \left(p_e^2c^2 + m_e^2c^4\right)^{1/2}.$$

Solving this system would yield

$$\lambda' - \lambda = \frac{h}{m_e c}(1 - \cos\theta)$$

which, again, is consistent with experiment. This phenomenon is known as Compton scattering.

At this point we're confident that light comes in discrete quanta, so we might wonder now how these quanta behave. We can investigate this using a Mach-Zehnder interferometer, illustrated at right.



Elements B and C are mirrors, while the semicircles are detectors. Elements A and D are beam splitters, which reflect half of any incident light and transmit the rest. If a single photon is incident, though, that photon doesn't get cut in half and take both paths; rather, it picks one path and sticks with it. This pick is random, too—there is no way to deterministically predict which way the photon will go. The best we can do is work in probabilities.

It turns out that these probabilities aren't as simple as flipping a coin for each splitter. Assuming photons are sent into splitter A at a constant rate, experiments show that the rate at which either of the detectors receives photons varies sinusoidally with distance. This suggests that the relevant quantity is not just a probability, but also an associated phase. Complex numbers happen to describe this behavior in just the right way!

In our theory of quantum optics, each event has an associated complex number z called a probability amplitude. There are three fundamental rules that these probability amplitudes abide by, and they're surprisingly intuitive.

1. If an event has probability amplitude z with complex conjugate z^*, then the probability of the event occurring is given by
$$P = |z|^2 = z^* z.$$

2. If an event can be broken down ito a series of steps, then its probability amplitude is the product of those for each step:
$$z = z_1 z_2 z_3 \cdots .$$

3. If an event can happen in multiple independent ways, then its probability amplitude is the sum of those for each way:
$$z = z_1 + z_2 + z_3 + \cdots .$$

Below is a table of probability amplitudes for some events that are relevant to our discussion.

Event	Amplitude
travel	$z = e^{ikd},\ k = 2\pi n/\lambda$
transmission	$z = \sqrt{P_{\text{trans}}}$
reflection	$z = \pm\sqrt{P_{\text{ref}}}$

Here d is the distance a photon travels in a medium with refractive index n. (Note that $n = 1$ for a vacuum.) The probability amplitude for photon reflection has a $\pm$ to account for the phase shift that occurs at interfaces with higher-index materials.

4.4 Multiple-slit Diffraction

Consider a laser that shines through two very narrow slits and onto a faraway wall. On the wall we may expect to see two bright peaks of light, but instead there's a row of bright bands that peak right between the slits. There are both classical and quantum explanations for this phenomenon; we will, of course, focus on the quantum one here.

There are two paths light can take to get to any given point on the wall, one for each slit. Let d be the small distance between the slits, d_0 the distance from the laser to the slits, and d_1, d_2 the distances from each slit to the wall. So we have the probability amplitudes
$$z_1 = e^{ikd_0} r e^{ikd_1} \quad \text{and} \quad z_2 = e^{ikd_0} r e^{ikd_2},$$
where r is the probability amplitude of the light going in the proper direction to reach the point on the wall. (Strictly speaking r depends on the deflection angle θ, which is different for the two slits. The difference is so slight, though, that they're approximately equal.) Adding these amplitudes gives
$$z = z_1 + z_2 = r e^{ikd_0} e^{ikd_1} \left(1 + e^{ik(d_2 - d_1)}\right),$$
and the probability
$$P = 4r^2 \cos^2\left(\frac{k(d_2 - d_1)}{2}\right).$$
If the wall is very far away then we have $d_2 - d_1 = d\sin\theta$ and
$$P = 4r^2 \cos^2\left(\frac{kd\sin\theta}{2}\right) = 4r^2 \cos^2\left(\frac{\pi d\sin\theta}{\lambda}\right).$$
Notice that this probability is maximized when the phase difference $kd\sin\theta$ is an integer, so that the each slit produces light perfectly in phase. Similarly, the probability is minimized when the phase difference is at midpoints between integers. Thus we see
$$\text{maxima at } d\sin\theta = 0\lambda,\ 1\lambda,\ 2\lambda,\ \ldots \quad \text{and} \quad \text{minima at } d\sin\theta = \frac{1}{2}\lambda,\ \frac{3}{2}\lambda,\ \frac{5}{3}\lambda,\ \ldots .$$
As we increase the number of slits, the width of each bright spot decreases to a point. This is because, as we increase the number of slits, there are more values of θ that can cause destructive interference. To see this, note how the probability for n slits is given by
$$P = r e^{ikd_0} e^{ikd_1} \left(1 + e^{i\theta} + e^{i(2\theta)} + \cdots + e^{i(n\theta)}\right).$$
We might model each slit's complex exponential as a phasor, and that all of these phasors are being summed to get the total probability for a point on the wall. If for a particular θ the phasors sum to zero (which manifests itself in phasor space as a regular n-gon), then there will be perfect destructive interference there.

4.5 Single-slit Diffraction

Now suppose that, instead of being faced with a row of infinitesimal slits, we have a single slit of finite width a that is parallel with a faraway wall. Suddenly, the path of an incident photon can no longer be entirely described by its angle of deflection!

Consider an infinitesimal bit dx of the slit, located a distance x from one of its edges. The probability amplitude associated with a photon passing through this bit and landing on a particular spot on the wall is

$$dz = dr\, e^{ik(d_1 + x\sin\theta)},$$

where d_1 is the distance between the edge of the slit and the spot on the wall, θ is the direction in which a photon must be deflected to hit this spot, and dr is the probability amplitude associated with this deflection.

The probability amplitude for the entire slit is the integral of dz over the entire slit. We'll take $dr = (r/a)dx$ in order to normalize our probabilities (where we've assumed that dr is independent of where we are on the slit), so we have

$$z = \int_0^a \frac{r}{a} dx\, e^{ik(d_1+\sin\theta)} = \frac{re^{ik(d_1+(a/2)\sin\theta)}}{ika\sin\theta}\left(e^{\frac{1}{2}ika\sin\theta} - e^{-\frac{1}{2}ika\sin\theta}\right) = \frac{re^{i(\cdots)}}{\frac{ka}{2}\sin\theta}\sin\left(\frac{ka}{2}\sin\theta\right).$$

Defining $\alpha = (ka/2)\sin\theta$ gives

$$P = \frac{r^2}{\alpha^2}\sin^2\alpha.$$

On our faraway wall we would see one bright peak at the center of the slit, and then a bunch of much dimmer peaks going off to the side. (About ninety percent of the intensity of the light is concentrated within this central bump.) We see perfect destructive interference at integer multiples of π for α, meaning the first minima $\pm\theta_1$ occur where $\sin\theta_1 = \lambda/a$.

This actually gives us a criterion for determining whether or not two objects, say headlights on a car, are distinguishable from a distance! Specifically, according to the Rayleigh criterion two objects are "barely resolvable" when the center of one diffraction pattern is at the minimum of the other. Any closer and they mesh together into one.

Now, in this entire discussion of quantum optics we've ignored one glaring issue: for any point in space, there's actually an infinite number of paths a photon can take to get there! There's no obvious reason light should prefer traveling in straight lines over, anything else. The principle of least time resolves this issue: if there is a continuous range of paths that light can take between two points in space, the only path that contributes significantly to the probability of detection at the end point is that of minimum time (or, more importantly, minimum phase). At a high level, this is because the shorter paths are only marginally different in phase and so tend to constructively interfere, while longer paths are quite different from one another and tend to cancel one another out.

Finally, we cap everything off with one more disturbing revelation. Everything we've just discussed–diffraction, interference, all that–also works with massive particles like helium atoms. This strongly suggests that matter, too, has wave-like properties that can be described in terms of probability amplitudes. We might make a leap in logic and define the (nonrelativistic) de Broglie wavelength

$$\lambda_{\mathrm{dB}} = \frac{h}{mv},$$

inspired by how $E = h\nu = pc$ for a photon, meaning it has $\lambda = h/p$. In principle every bit of matter has a de Broglie wavelength, but at the macroscopic level the wavelengths are far too small relative to the apertures we face in real life to cause any diffraction. Despite that oddity, this leap is a good one, and it turns out to be exactly what we need to facilitate a proper study of quantum mechanics.