Scott Hughes *1 February 2005*

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 1:
Introduction; Couloumb's law; Superposition; Electric energy

1.1 Fundamental forces

The electromagnetic force is one of the four fundamental forces of physics. Let's compare it to the others to see where it fits in the grand scheme of things:

Force	Range (cm)	Interaction particles	Exchange particles	"Strength"
Gravity	∞	All mass/energy	graviton ($m = 0$)	1
Weak	10^{-15}	All elementary particles except photon & gluons	weak bosons ($m \sim 100 \times m_{\rm proton}$)	10^{24}
Electromagnetic	∞	All charges	photon ($m = 0$)	10^{35}
Strong	10^{-13}	quarks and gluons	gluons ($m = 0$)	10^{37}

[The "strength" entries in this table reflect the relative magnitudes of the various forces as they act on a pair of protons in an atomic nucleus. Note that MIT professor Frank Wilczek shared the 2004 physics Nobel Prize with David Gross (now director of the Kavli Institute for Theoretical Physics at UC Santa Barbara) and H. David Politzer (now professor of physics at Caltech) for work that made possible our current understanding of the strong force.]

Let's summarize the major characteristics of the electromagnetic force:

- **Strong**: It is the 2nd strongest of the fundamental forces — only the appropriately named strong force (which binds quarks and nuclei together) beats it (and only at very small distances).
- **Long Range**: As we'll discuss soon, the electromagnetic force has the form $F \propto 1/r^2$, just like gravity. This means it has *infinite range*: it gets weaker with distance, but just slowly enough that things really far away still feel it. Contrast this with the weak force, which has the form $F \propto \exp(-r/R_{\rm W})/r^2$, where $R_{\rm W} \simeq 10^{-15}$ cm — it dies away REALLY fast with r.
- **Has 2 signs**: It can attract and repel. Contrast this with gravity which only attracts.
- **Acts on charge**: More on this shortly.

The electromagnetic interaction plays an *extremely* important role in everyday life. The forces that govern interactions between molecules in chemistry and biology (e.g., ionic forces, van der Waals, hydrogen bonding) are all fundamentally electromagnetic in nature. A thorough understanding of electromagnetism will carry you far in understanding the properties of things you encounter in the everyday world. Add quantum mechanics and you've got most of the tools needed to completely understand these things.

1.2 History of electromagnetism

Circa B.C. 500: Greeks discover that rubbed amber attracts small pieces of stuff. (Note the Greek word for amber: "$\eta\lambda\epsilon\kappa\tau\rho o\nu$", or "electron".) They also discover that certain iron rich rocks from the region of $\mathrm{M}\alpha\gamma\nu\eta\sigma\iota\alpha$ (Magnesia) attract other pieces of iron.

1730: Charles Francois du Fay noted that electrification seemed to come in two "flavors": *vitreous* (now known as positive; example, rubbing glass with silk) and *resinous* (negative; rubbing resin with fur)

1740: Ben Franklin suggested the "one fluid" hypothesis: "positive" things have more charge than "negative" things. Suggested an experiment to Priestley to *indirectly* measure the inverse square law...

1766: Joseph Priestley: did the suggested experiment, indirectly proving $1/r^2$ form of the force law.

1773: Henry Cavendish: did Priestley's experiment accurately:

$$F = k\frac{q_1 q_1}{r^{2+\delta}}\ , \qquad |\delta| < \frac{1}{50}\ .$$

(By the late 1800s, Maxwell was able to show that $|\delta| < 1/21600$; by 1936 Plimpton and Lawton were able to show $|\delta| < 2 \times 10^{-9}$.)

1786: Charles-Augustin de Coulomb measured the electrostatic force and *directly* verified the inverse square law.

1800: Count Alessandro Giuseppe Antonio Anastasio Volta invented the electric battery.

1820: Hans Chrstian Oersted and André-Marie Ampère establish connection between magnetic fields and electric currents

1831: Michael Faraday discovers magnetic induction

1873: James Clerk Maxwell unified electricity and magnetism into electromagnetism.

1887: Heinrich Hertz confirms the connection between electromagnetism and radiation.

1905: Albert Einstein formulates the special theory of relativity, which (among other things) clarifies the inter-relationship between electric and magnetic fields.

1.3 Electric charge

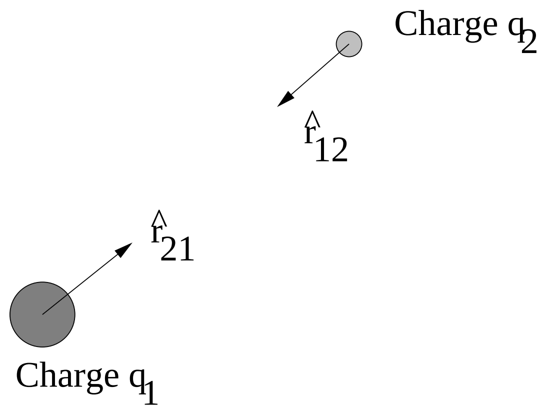
As stated above, the electromagnetic interaction acts on *charges*. This is really a tautological statement: all we are doing is defining "charge" as the property that the electromagnetic interaction interacts with. We accept this and move on.

In nature, we find that electric charge comes in discrete lumps or "quanta". This was first measured by Robert Millikan in 1923 in his famous oil drop experiment. The smallest charge we ever observe is the "elementary charge" e; it takes the value 1.602×10^{-19} Coulombs (one of the units of charge we will use), or 4.803×10^{-10} esu (electrostatic units). The magnitude of the electron's charge and of the proton's charge is e. The charges we encounter in nature come in opposite signs. By convention, the electron has negative sign, the proton positive.

Charge is conserved: in any isolated system, the total charge cannot change. If it does change, then the system is not isolated: charge either went somewhere or came in from somewhere.

1.4 Coulomb's law

Coulomb's law tells us that the force between two charges is proportional to the product of the two charges, inversely proportional to the distance between them, and points along the vector pointing from one charge to the other:



The unit vector $\hat{r}_{21}$ points *toward* 2 *from* 1; $\hat{r}_{12}$ points *toward* 1 *from* 2. This means that $\hat{r}_{12} = -\hat{r}_{21}$. We write the distance to charge 2 from 1 as r_{21}; this must be exactly the same as the distance to charge 1 from 2, r_{12}.

According to Coulomb's law, the force that charge q_2 feels due to charge q_1 is given by

$$\vec{F}_2 = k\frac{q_1 q_2}{r_{21}^2}\hat{r}_{21}$$

(We'll discuss the constant k shortly. The key thing to note right now is that it is *positive.*) Likewise, the force q_1 feels due to charge q_2 is

$$\begin{aligned}\vec{F}_1 &= k\frac{q_1 q_2}{r_{12}^2}\hat{r}_{12} \\ &= -k\frac{q_1 q_2}{r_{12}^2}\hat{r}_{21} \\ &= -\vec{F}_2\ .\end{aligned}$$

The force on q_1 is equal but opposite to the force on q_2. Newton would be pleased (3rd law).

It's worth noting that we are *assuming* Coulomb's law. Of course, this assumption is justified because many many experiments have verified that this is the way the electric force works. It's kind of interesting, though, that pretty much this entire class follows from this assumption. [Indeed, there are only three other "fundamental" assumptions that we will need — superposition (discussed momentarily), special relativity, and the non-existence of magnetic charge. All of 8.022 can be derived from these starting points!]

To illustrate Coulomb's law, we have introduced all these subscripts and made it seem like half of our time is going to be spent labeling charges, unit vectors, and distances. In practice, though, the thing to note is that the force *in both cases* is along the vector from one charge to the other. Its direction in both cases is *repulsive* if both charges have the same sign; it is *attractive* if the charges are of opposite sign. Like charges repel; opposites attract.

1.5 Units

In electrodynamics, the choice of units we use matters: the basic equations actually have slightly different forms in different system of units. Contrast this with mechanics, where $\vec{F} = m\vec{a}$ regardless of whether we measure length in units of meters, inches, furlongs, or lightyears.

The system of units that is used by our textbook, and that is common in much physics research (particularly theory) is the *CGS* unit system: length is measured in centimeters, masses in grams, time in seconds. In this system of units the constant k appearing in Coulomb's law is equal to 1. The cgs unit of charge is the *esu*, an acronym for the oh-so-clever name *e*lectro*s*tatic *u*nit. It is defined such that two charges of 1 esu each, separated by 1 cm, feel a force of 1 dyne:

$$1\,\mathrm{dyne} = \frac{(1\,\mathrm{esu})^2}{(1\,\mathrm{cm})^2} \rightarrow 1\,\mathrm{esu} = \mathrm{cm}\,\sqrt{\mathrm{dyne}}\,.$$

Another system of units that is commonly encountered is known as *SI* (for Systèm Internationale) or *MKS* units. Engineers typically use SI units in their work. Lengths are measured in meters, masses in kilograms, time in seconds. Charges are measured in *Coulombs*, abbreviated C. In this system of units the constant k in Coulomb's law takes the value

$$k = \frac{1}{4\pi\epsilon_0}$$

where ϵ_0 is yet another constant called the permittivity of free space. It takes the value

$$\epsilon_0 = 8.854 \times 10^{-12}\,\mathrm{C}^2\,\mathrm{N}^{-1}\,\mathrm{m}^{-2}\,.$$

The constant k then takes the value

$$\begin{aligned} k &= 8.988 \times 10^9\,N\,\mathrm{C}^{-2}\mathrm{m}^2 \\ &\simeq 9 \times 10^9\,N\,\mathrm{C}^{-2}\mathrm{m}^2 \end{aligned}$$

Note that the force between two 1 Coulumb charges held a meter apart is 9 *billion* Newtons! 1 Coulomb is an enormous amount of charge: 1 Coulomb = 2.998×10^9 esu. (The 2.998 is often approximated 3. Note that the speed of light's value is 2.998×10^{10} cm/sec. The equivalence of the numerical factors is *not* a coincidence!)

A common complaint in this class is that "no one uses cgs units, so why should we"? This complaint is not really valid — people do in fact use cgs for a *lot* of things, as well as even more bizarre variants of cgs units. (For example, in my research I work in a version of cgs units in which both time and mass are measured in centimeters. In this system, the speed of light c and the gravitational constant G are both equal to exactly 1. In some cases, people actually *mix* units. The most commonly used unit for magnetic field is the Gauss, which is cgs — even if everything else is measured in SI. Never count on people to adopt a logical system of anything!)

It is true, however, that cgs units are not encountered as often as SI units, so I understand this annoyance. The harsh truth is that the best thing to do is to adapt, learn how to convert between different unit conventions, and deal with it. It's actually good practice for being a scientist or engineer — you'd be stunned how often people use what might be considered nonstandard choices for things, especially if you have to dig back into older documents.

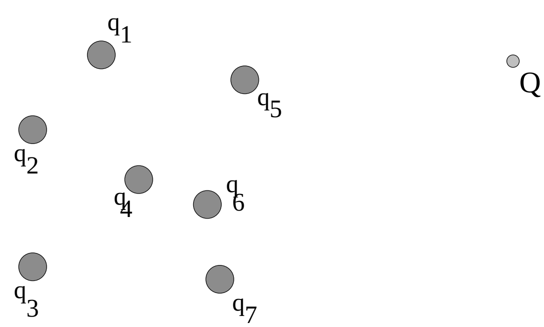
Here is a summary of how to convert various quantities between SI and CGS. In this table, "3" is shorthand for 2.998. (More precisely, 2.99792458). "9" means "3" squared, or 8.988. (More precisely, 8.9875517873681764.) In the vast majority of calculations, approximating "3" as 3 and "9" as 9 is plenty accurate.

	SI Units		CGS units
Energy	1 Joule	=	10^7 erg
Force	1 Newton	=	10^5 dyne
Charge	1 Coulomb	=	"3" $\times 10^9$ esu
Current	1 Ampere	=	"3" $\times 10^9$ esu/sec
Potential	"3" $\times 10^2$ Volts	=	1 statvolt
Electric field	"3" $\times 10^4$ Volts/m	=	1 statvolt/cm
Magnetic field	1 Tesla	=	10^4 gauss
Capacitance	1 Farad	=	"9" $\times 10^{11}$ cm
Resistance	"9" $\times 10^{11}$ Ohm	=	1 sec/cm
Inductance	"9" $\times 10^{11}$ Henry	=	1 sec^2/cm

1.6 Superposition

The principle of superposition is a fancy name for a simple concept: in electromagnetism, forces add.

More concretely: suppose we have a number N of charges scattered in some region. We want to calculate the force that *all* of these charges exert on some "test charge" Q. Label

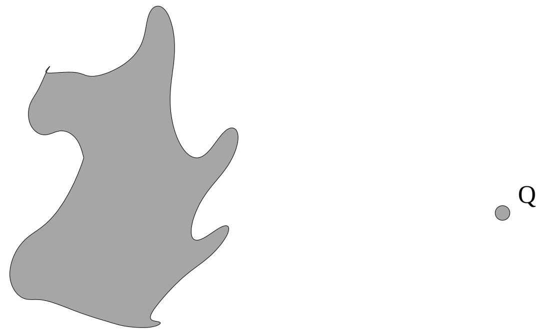


the ith charge q_i; let r_i be the distance from q_i to Q. Let $\hat{r}_i$ be the unit vector pointing from q_i to Q. The total force acting on Q is just the total found by summing the forces from each *individual* charge q_i:

$$
\begin{aligned}
\vec{F}_Q &= \frac{Qq_1\hat{r}_1}{r_1^2} + \frac{Qq_2\hat{r}_2}{r_2^2} + \frac{Qq_3\hat{r}_3}{r_3^2} + \ldots \\
&= \sum_{i=1}^{N} \frac{Qq_i\hat{r}_i}{r_i^2} \,.
\end{aligned}
$$

Since we know calculus, we can take a limiting description of this calculation. Let's imagine that, in our swarm of charges, the number N becomes infinitely large, the charges all press together, and the whole thing goes over to a *continuum*. We describe this continuum by the *charge density* or charge per unit volume ρ.

Imagine that we've got a big blob of charged goo with density ρ sitting somewhere. We bring in our test charge Q close to it. How do we calculate the total force acting on the test charge?



Simple: we chop the blob up into little chunks of volume ΔV; each such chunk contains charge $\Delta q = \rho \Delta V$. Suppose there are N total such chunks, and we label each one with some index i. Let $\hat{r}_i$ be the unit vector pointing from the ith chunk to the test charge; let r_i be the distance between chunk and test charge. The total force acting on the test charge is

$$\vec{F} = \sum_{i=1}^{N} \frac{Q(\rho \Delta V_i)\hat{r}_i}{r_i^2}$$

This formula is really an approximation, since we probably can't perfectly describe our blob by a finite number of little chunks. The approximation becomes *exact* if we let the number of chunks go to infinity and the volume of each chunk go to zero — the sum then becomes an integral:

$$\vec{F} = \int_V \frac{Q\,\rho\,dV\,\hat{r}}{r^2} \ .$$

If the charge is uniformly smeared over a surface, then we integrate a *surface* charge density σ over the area of the surface:

$$\vec{F} = \int_A \frac{Q\,\sigma\,dA\,\hat{r}}{r^2} \ .$$

If the charge is uniformly smeared over a line, then we integrate a *line* charge density λ over the length:

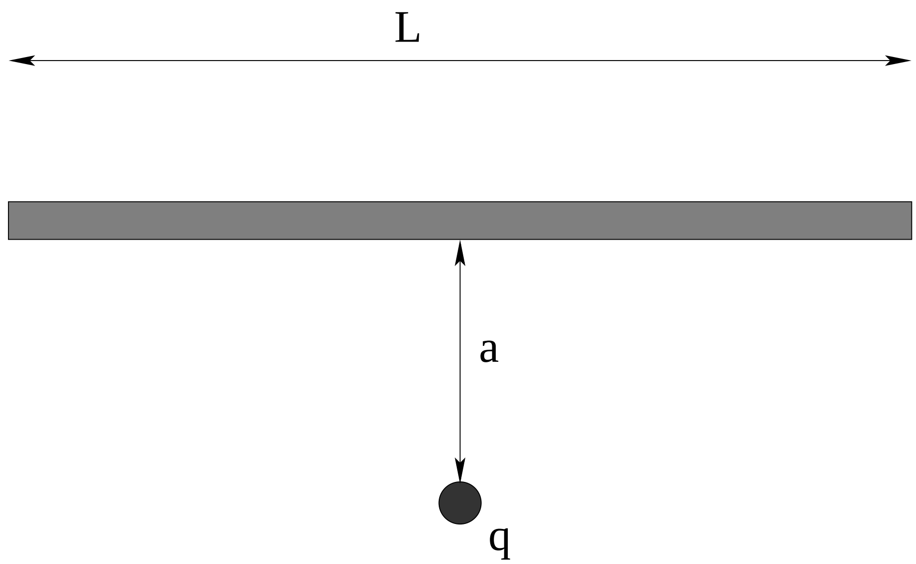
$$\vec{F} = \int_L \frac{Q\,\lambda\,dl\,\hat{r}}{r^2} \ .$$

(Note that the letter σ is the Greek equivalent for s, so it makes sense for a surface distribution. λ is the Greek equivalent for l, so it makes sense for a line distribution. ρ is the Greek equivalent for r so it makes sense ... actually, it makes no sense at all. It was almost logical.)

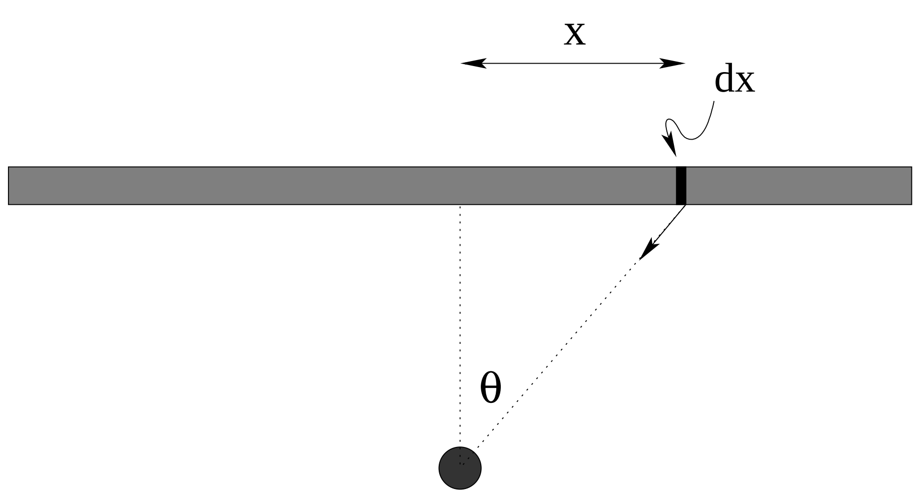
A good portion of the calculations that we will do will be like this. Let's look at an example:

1.6.1 Example: Force on a test charge q due to a uniform line charge

A rod of length L has a total charge Q smeared uniformly over it. A test charge q sits a distance a away from the rod's midpoint. (Ignore the thickness of the rod.) What is the electric force that the rod exerts on the test charge?



First, define a coordinate system: let the direction parallel to the rod be x, let the perpendicular direction be y. Next, imagine dividing the rod up into little pieces of width dx:



In drawing things this way, we have defined $x = 0$ as the rod's midpoint, so the rod runs from $x = -L/2$ to $x = L/2$. We have also drawn the unit vector $\hat{r}$ that points from the piece of the rod to the test charge. *Note that the unit vector's direction changes as we slide the piece dx from one end of the rod to the other.* This is a *very* important thing to note: the unit vector $\hat{r}$ is a *function*, not just a constant. With trigonometry, it is simple to work out what this function is:

$$\hat{r} = -\hat{x}\sin\theta - \hat{y}\cos\theta\ .$$

With a bit more trigonometry, we can write this as a function of x, the location of the chunk, rather than of θ:

$$\hat{r} = -\hat{x}\frac{x}{\sqrt{a^2 + x^2}} - \hat{y}\frac{a}{\sqrt{a^2 + x^2}}\ .$$

Notice finally that $\sqrt{a^2 + x^2}$ is the distance from the element dx to the test charge.

Defining $\lambda = Q/L$, the force that the test charge feels is given by a simple integral:

$$\vec{F} = -\int_{-L/2}^{L/2} q\lambda \frac{x\hat{x} + a\hat{y}}{(a^2 + x^2)^{3/2}}\, dx \ .$$

The horizontal ($\hat{x}$) component is obviously zero: for every element on the right of the midpoint, there is an element on the left whose force magnitude is equal, but whose horizontal component points in the opposite direction. You can prove this by doing the integral, but it's very healthy to understand the physics of this simple *symmetry argument*.

The remaining integral for the y component is now easy: using

$$\int \frac{dx}{(a^2 + x^2)^{3/2}} = \frac{x}{a^2\sqrt{a^2 + x^2}} \ ,$$

we find

$$\begin{aligned} \vec{F} &= -\frac{q\lambda L}{a\sqrt{a^2 + L^2/4}}\hat{y} \\ &= -\frac{qQ}{a\sqrt{a^2 + L^2/4}}\hat{y} \ . \end{aligned}$$

This is a typical example of the kinds of integrations we will need to do (though most will be more challenging than this simple case).

1.7 Energy of a system of charges

1.7.1 Work done moving charges

Suppose we have a charge Q sitting at the origin. How much work must we do to move a charge q from radius r_2 to radius r_1? Let us first do this assuming that q moves on a purely radial path. The work that we do is given by integrating the force that *we exert* along this path:

$$W = \int \vec{F}_{\rm us} \cdot d\vec{s}$$

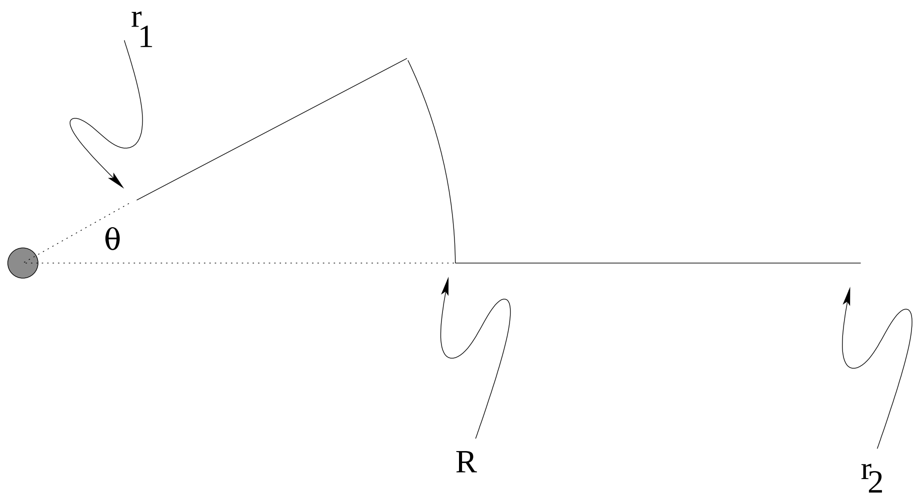
What is $F_{\rm us}$? It is *opposite* to the force that the charge Q exerts on q:

$$\vec{F}_{\rm us} = -\vec{F}_{\rm Coulomb} = -\frac{Qq\hat{r}}{r^2} \ .$$

We integrate this along a radial path, $ds = dr\,\hat{r}$, from $r = r_2$ to $r = r_1$:

$$\begin{aligned} W(r_2 \to r_1) &= \int \vec{F}_{\rm us} \cdot d\vec{s} \\ &= -\int_{r_2}^{r_1} \frac{Qq}{r^2} dr \\ &= \frac{Qq}{r_1} - \frac{Qq}{r_2} \ . \end{aligned}$$

It is straightforward to prove that this result holds along *any* path. Suppose we come in on a radial path to radius R, move in a circular arc through some angle θ, and then continue on a new radial path to r_1:



The differential $d\vec{s}$ in this case is still $dr\,\hat{r}$ along the two radial paths, but it is $R\,d\theta\,\hat{\theta}$ along the angular segment (where $\hat{\theta}$ is a unit vector that points in the angular direction). The integral thus becomes

$$\begin{aligned} W(r_2 \to r_1) &= \int \vec{F}_{\rm us} \cdot d\vec{s} \\ &= -\int_{r_2}^{R} \frac{Qq}{r^2} dr - \int_0^{\theta} \frac{Qq}{R^2} (\hat{r} \cdot \hat{\theta})(R\,d\theta) - \int_R^{r_1} \frac{Qq}{r^2} dr \\ &= \frac{Qq}{R} - \frac{Qq}{r_2} + 0 + \frac{Qq}{r_1} - \frac{Qq}{R} = \frac{Qq}{r_1} - \frac{Qq}{r_2} \;. \end{aligned}$$

We end up with the *exact same result* because the contribution around the arc is zero: that path is perpendicular to the force, so we do no work moving along that segment.

A corollary of this result is that the work we do is the same along *ANY* path that we take from r_2 to r_1. Why? We can break up any path into radial segments and angular arcs, in a similar manner to the above calculation. The angular segments don't matter; the radial segments will give us the same answers we've just derived.

A second corollary is that the electric force is *conservative*: any path that takes us from r_1 back to r_1 does zero work (just plug $r_2 = r_1$ into the above result).

1.7.2 Work done to assemble a system of charges

How much work does it take to *assemble* a system of charges? With the calculations we just did, answering this is easy.

Let's clarify the question a bit: how much work does it take to move a charge q_2 in from infinity until it is a distance r_{12} from a charge q_1? This is *really* what we mean what we ask how much energy it takes to assemble the system — we are asking what it takes to take charges that are very far apart and bring them close together. Using the results we worked out above, the answer is clearly

$$\begin{aligned} W(\infty \to r_{12}) &= \frac{q_1 q_1}{r_{12}} \\ &\equiv W_{12} \;. \end{aligned}$$

This quantity can then be thought of as the energy content U of this system of charges: $U = W$.

Now, how about if we bring in a third charge q_3? This turns out to be easy using the principle of superposition. To make it very clear what is going on, let's go back to the integral we have to do for the work: we bring the charge q_3 in from infinite distance until it is a distance r_{13} from charge q_1 and r_{23} from charge q_2:

$$\begin{aligned} W_{(1+2)3} &= -\int_\infty^{r_{13}} \frac{q_1 q_3}{r^2} dr - \int_\infty^{r_{23}} \frac{q_2 q_3}{r^2} dr \\ &= \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}} \\ &= W_{13} + W_{23} \; . \end{aligned}$$

We add this to the work it took to bring q_1 and q_2 together to get the total work to assemble this system:

$$\begin{aligned} W &= W_{12} + W_{13} + W_{23} \\ &= \frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}} \; . \end{aligned}$$

Again, this is the total energy content of the system: $U = W$.

We can generalize this to as many charges as we want; the resulting formulas rapidly get pretty big. The key thing to note is that we are summing over *pairs* of charges in the system. It is pretty simple to write down a formula that sums up the energy over pairs like this:

$$U = \frac{1}{2} \sum_{j=1}^{N} \sum_{\substack{k=1 \\ j \neq k}}^{N} \frac{q_j q_k}{r_{jk}}$$

The requirement that $j \neq k$ makes sure that we only count over pairs of charges — if $k = j$, we get nonsense in the formula. The factor of $1/2$ out front is needed because the sums will actually count each pair twice. (Try it out for a simple case, $N = 2$ or $N = 3$, to see.)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 2:
ELECTRIC FIELDS & FLUX; GAUSS'S LAW

2.1 The electric field

Last time, we learned that Coulomb's law plus the principle of superposition allowed us to write down fairly simple formulas for the force that a "swarm" of N charges exerts on a test charge Q:

$$\vec{F} = \sum_{i=1}^{N} \frac{Q\, q_i\, \hat{r}_i}{r_i^2} \ .$$

We found another simple formula for the force that a blob of charged material with charge density ρ exerts on a test charge:

$$\vec{F} = \int \frac{Q\, \rho\, dV\, \hat{r}_i}{r_i^2} \ .$$

And, of course, similar formulas exist if the charge is smeared out over a surface with surface density σ, or over a line with line density λ.

In all of these cases, the force ends up proportional to the test charge Q. We might as well factor it out. Doing so, we define the *electric field*:

$$\begin{aligned}
\vec{E} &\equiv \frac{\vec{F}}{Q} \\
&= \sum_{i=1}^{N} \frac{q_i\, \hat{r}_i}{r_i^2} \qquad (N \text{ Point charges}) \\
&= \int \frac{\rho\, dV\, \hat{r}_i}{r_i^2} \qquad (\text{Charge continuum}) \ .
\end{aligned}$$

Given an electric field, we can figure out how much force will be exerted on some point charge q in the obvious way: $\vec{F} = q\vec{E}$.

In cgs units, the electric field is measured in units of dynes/esu, or esu/cm^2, or statvolts/cm — all of these choices are equivalent. (We'll learn about statvolts in about a week.) In SI units, first of all, we must multiply all of the above formulae by $1/4\pi\epsilon_0$. The units of electric field are then given in Newtons/Coulomb, or Coulomb/meter^2, or Volts/meter.

Let's look at a couple of examples.

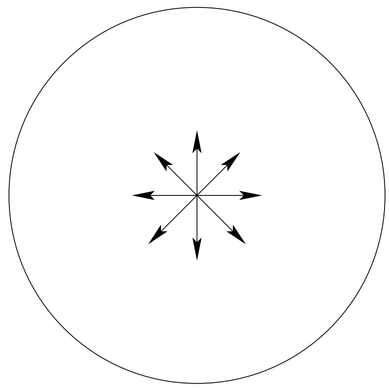
2.1.1 Example: Electric field at center of charged ring

Take a ring of radius R with uniform charge per unit length $\lambda = Q/2\pi R$. What is the electric field right at its center? Well, if we don't think about it too carefully, we want to do something like this:

$$\vec{E} = \int \frac{\lambda\, \hat{r}}{r^2} ds \,,$$

where the integral will be take around the ring; ds will be a path length element on the ring, r is the radius of the ring, and $\hat{r}$ is a unit vector pointing from each length element on the ring to the ring's center ...

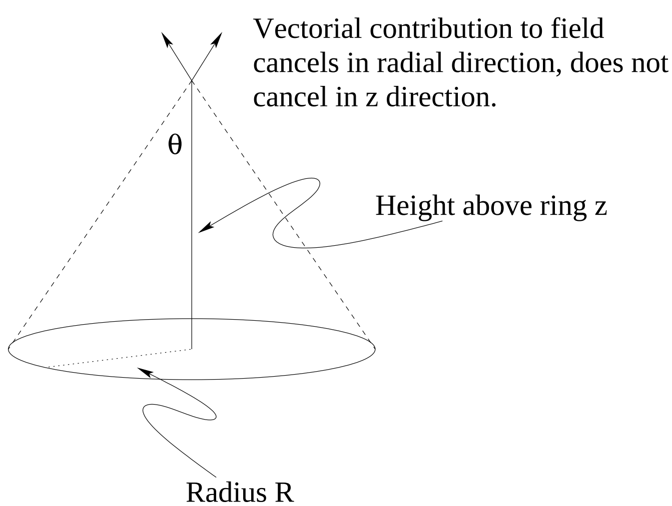
But wait a minute! You should quickly notice that for each contribution due to a little length element on the ring, there is an equal contribution from the ring's opposite side *pointing in the opposite direction*. Each contribution to the $\vec{E}$ field at the center of the ring is precisely canceled by its opposite contribution:



So the field at the center of the ring is zero.

2.1.2 Example: Electric field on ring's axis

Suppose we move along the ring's axis to some coordinate z above the disk's center. The field does not cancel in this case, at least not entirely: Considering the contribution of pieces of the ring, we see that there is a component in the radial direction, and a component along z. The radial component cancels by the same symmetry argument as we used above. The z component, however, definitely does not cancel:



So, let's calculate: the contribution coming from a little segment of the ring is

$$d\vec{E} = \frac{dQ}{r^2}\cos\theta\,\hat{z}$$

$$
\begin{aligned}
&= \frac{\lambda\, ds}{R^2 + z^2} \cos\theta\, \hat{z} \\
&= \frac{\lambda\, z\, ds}{(R^2 + z^2)^{3/2}}\, \hat{z}\,.
\end{aligned}
$$

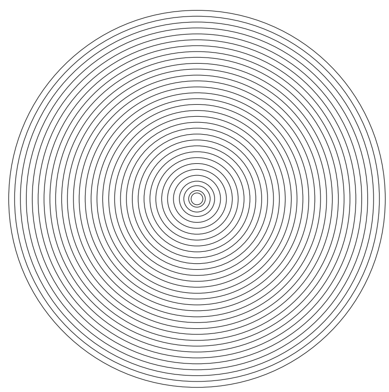
The length ds represents the length of a little segment of the ring. Integrating this up is trivial: we go around the circumference, so the only "action" as far as the integral is concerned is that we replace ds with $2\pi R$. We end up with

$$
\begin{aligned}
\vec{E} &= \frac{2\pi R\, \lambda\, z}{(R^2 + z^2)^{3/2}}\, \hat{z} \\
&= \frac{Q\, z}{(R^2 + z^2)^{3/2}}\, \hat{z}\,.
\end{aligned}
$$

2.1.3 Example: Electric field on disk's axis

Using that result, it is now simple to work out a more complicated $\vec{E}$ field, that above a disk. Suppose we have a charge Q smeared out uniformly over a disk of radius R, so that the charge per unit area $\sigma = Q/\pi R^2$. What is $\vec{E}$ a distance z above the disk?

A simple way to do this is to consider the disk to be made out of an infinite number of infinitesimally thin rings:



A single ring of radius r and thickness dr carries a total charge $dQ = \sigma\, 2\pi r\, dr$. The contribution of that ring to the disk's electric field can be easily figured using our ring electric field result, replacing R with r and Q with dQ:

$$
d\vec{E} = \sigma z \hat{z} \frac{2\pi r\, dr}{(z^2 + r^2)^{3/2}}\,.
$$

To get the total field, we integrate this from $r = 0$ to $r = R$:

$$
\begin{aligned}
\vec{E} &= \int_0^R \sigma z \hat{z} \frac{2\pi r\, dr}{(z^2 + r^2)^{3/2}} \\
&= 2\pi\sigma\hat{z} \left(\frac{z}{\sqrt{z^2}} - \frac{z}{\sqrt{R^2 + z^2}} \right)\,.
\end{aligned}
$$

Before moving on, it's useful to consider two limits of this formula. First, consider z large and positive: this formula can be rewritten

$$\begin{aligned}
\vec{E} &= 2\pi\sigma\hat{z}\left(1 - \frac{1}{\sqrt{1+(R/z)^2}}\right) \\
&\simeq 2\pi\sigma\hat{z}\left[1 - \left(1 - \frac{1}{2}\frac{R^2}{z^2}\right)\right] \\
&= \frac{\pi R^2\sigma\hat{z}}{z^2} = \frac{Q}{z^2}\hat{z}\ .
\end{aligned}$$

On the second line we have used $z \gg R$ to expand the square root via the binomial expansion. The final result tells us that, at large z the disk has the same $\vec{E}$ field as a point charge — which should not be a surprise!

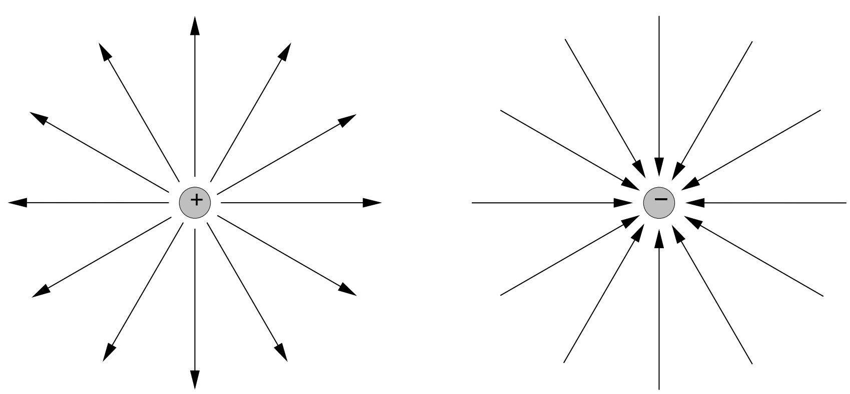
Next consider z very small and not necessarily positive: putting $\sqrt{z^2} = |z|$, we have

$$\begin{aligned}
\vec{E} &= 2\pi\sigma\hat{z}\left(\frac{z}{|z|} - \frac{z}{\sqrt{R^2+z^2}}\right) \\
&\simeq 2\pi\sigma\hat{z}\frac{z}{|z|}\ .
\end{aligned}$$

The function $z/|z| = +1$ above the disk, and -1 below the disk. Thus, our electric field has constant magnitude, $E = 2\pi\sigma$, and points *away* from the disk: up on the top side, down on the bottom. Remember this result — we will see it again soon!

2.2 Electric field lines

A useful concept to illustrate the electric field is the notion of *field lines*. These are fictitious lines we sketch which point in the direction of the electric field, and whose density gives us (at least roughly) an idea of how strong the field is. Here is a sketch of the electric field lines around a pair of point charges:

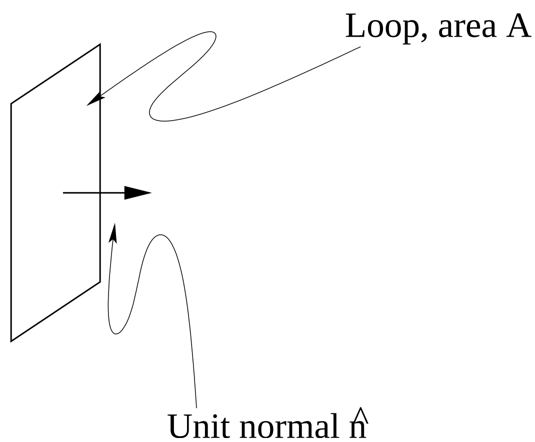


The field lines are bunched up near the charges (where the field is strongest), and spread out farther away. They point in towards the minus charge, and away from the plus charge.

Field lines can *never cross*. If they did, what would the field direction be at the crossing point? It wouldn't be defined! Sometimes you have situation where it *seems* like field lines are crossing; it usually turns out the magnitude goes to zero when that happens.

2.3 Flux

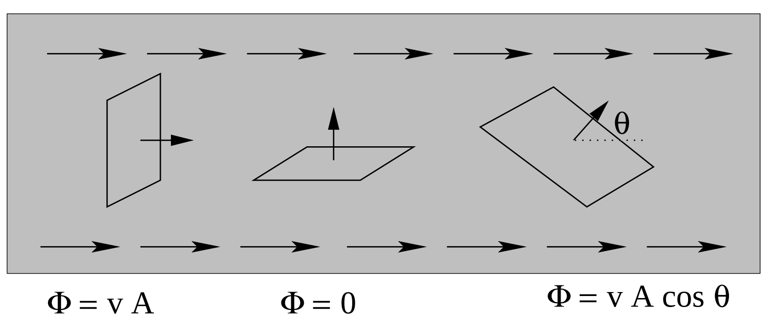
The notion of flux is an important one in physics. It refers to the *flow* of some vectorial quantity through an area. The simplest example to picture is the flow of fluid. Imagine that fluid is flowing with velocity $\vec{v} = v\hat{x}$. Now imagine that we dip a square wire loop of cross section area A into this fluid. We describe this loop with a vector $\vec{A} = A\hat{n}$, where $\hat{n}$ is the *normal* to the plane of the loop:



The velocity flux Φ_v is defined as the overlap between the velocity vector $\vec{v}$ and the loop area $\vec{A}$:

$$\Phi_v = \vec{v} \cdot \vec{A} \, .$$

For a uniform, plane loop, and constant, uniform velocity $\vec{v}$, we have $\Phi_v = vA\cos\theta$, where θ is the angle between $\hat{n}$ and the velocity vector:



For this example, the rate of fluid flow through this loop is proportional to this flux, Φ_v.

What if things aren't nice and uniform like this? Then, we break the cross sectional area of the loop up into little squares $d\vec{A}$; we assign each square its own normal $\hat{n}$. The flux then becomes an integral:

$$\Phi_v = \int \vec{v} \cdot d\vec{A} \, .$$

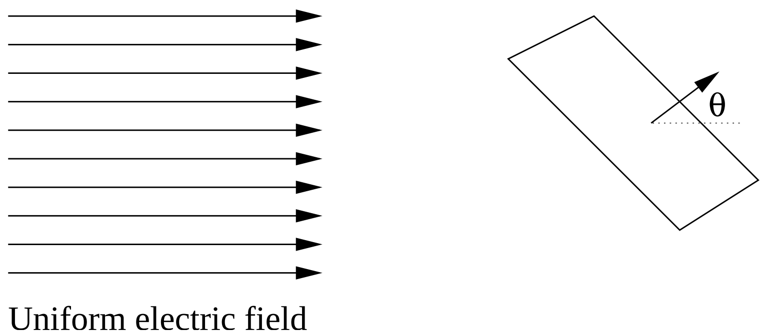
The integral is taken over the entire surface through which we wish to compute the flux.

2.4 Electric flux and Gauss's law (intuitive picture)

The electric flux Φ_E (which I'll usually just write Φ) is just like the velocity flux discussed above, using the electric field $\vec{E}$ rather than the fluid velocity $\vec{v}$:

$$\Phi = \int_{\mathrm{Surface}} \vec{E} \cdot d\vec{A} \; .$$

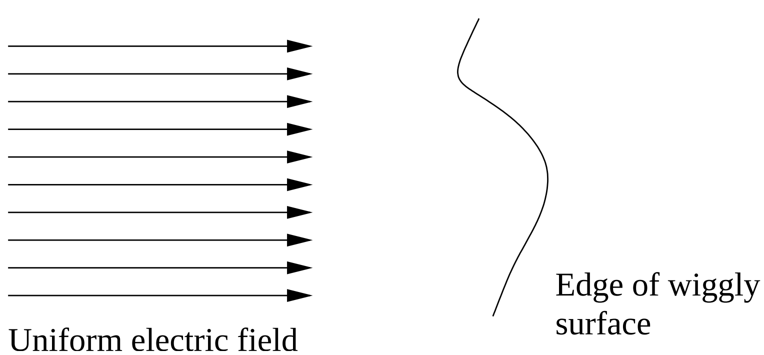
Electric flux has a very nice interpretation in terms of field lines. To motivate this interpretation, imagine that we have a uniform electric field. Let's calculate the flux through a square with cross sectional area A:



$$\Phi = (\vec{E}) \cdot (A\hat{n}) = EA\cos\theta \; .$$

No surprise — it looks just like the velocity flux we discussed above. The cool thing is how we can interpret this: the electric flux Φ is *proportional to* the number of field lines that pass through the loop![1] In this case, we can see that when $\theta = 0$ we have the maximum possible number of field lines passing through the loop. The number of lines that can fit decreases as we tilt the loop until, when $\theta = \pi/2$, *no* lines go through.

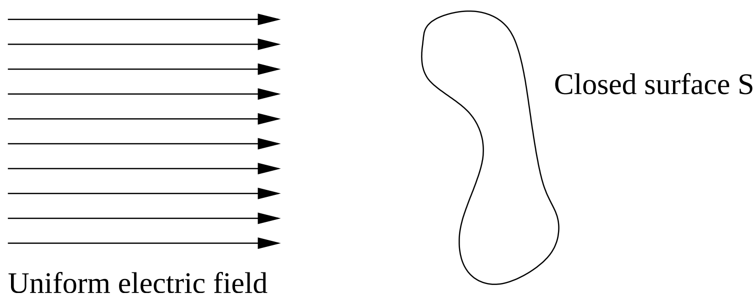
This interpretation holds perfectly well if we make our surface more complicated: if we make the surface wiggly,



[1] A reasonable objection to raise at this point is that field lines are really fictitious — is the notion that the number that passes through the loop at all meaningful? The answer is yes, provided that we remember it is just the proportionality that means anything. We can't actually calculate the *number* of field lines passing through the loop without defining some convention for what the density of field lines means. Clearly, though, we can see when the number that flow through is maximum — $\theta = 0$ — and minimum — $\theta = \pi/2$. Those statements are independent of any convention defining the field lines.

the total electric flux found by integrating over this surface still has this interpretation that it is the number of field lines passing through the surface.

A particular interesting and important case is when the surface is *closed*:

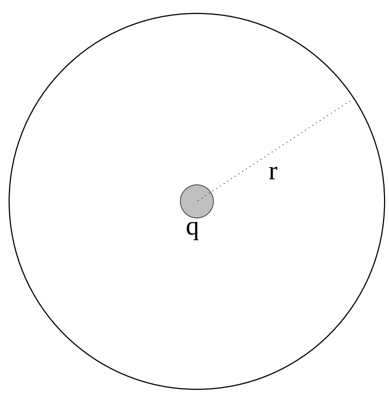


What is the flux that we find integrating over the total surface? With our intuitive interpretation, the answer is obvious:

$$\Phi = 0 \ .$$

Why? *Just as many field lines enter the surface as leave it!* Mathematically, you can see where this comes from by noting that the unit normal $\hat{n}$ for each area element $d\vec{A}$ is continually pointing in different directions as we move over the closed surface. On the left side of S, it mostly points to the left; on the right, mostly to the right. Summing over all elements we end up with a net cancellation.

What if we have a closed surface that only has field lines entering or leaving? As an example, imagine that we enclose a point charge q within a spherical surface of radius r:

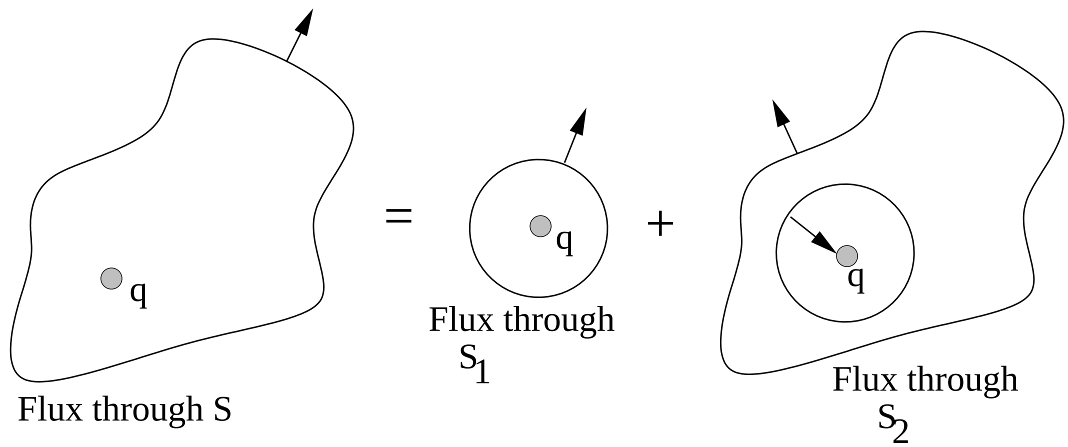


The electric flux through this surface is

$$\Phi = \oint \vec{E} \cdot d\vec{A} = \frac{q\hat{r}}{r^2} \cdot 4\pi r^2 \hat{r} = 4\pi q \ .$$

(The symbol $\oint$ is used for integrals over closed surfaces, or around a closed path, etc.) The integral was simple to do: since the surface of the sphere is at radius r, the $\vec{E}$ field is constant there; and, the normal vector is just the radial unit vector $\hat{r}$.

What if we the surface is not spherical? It turns out we get exactly the same answer! I'll derive this carefully in the next section. An intuitive explanation is to use the *Gaussian amoeba*. We take an arbitrarily shaped closed surface S containing a charge. We then carve a sphere out of it, breaking the region up into the "nucleus" and "protoplasm" of our amoeba:



Notice, in this picture, the orientation of the normals to the various surfaces. The original surface S has its normals pointing outward, as usual. Normals to the bounding surface of the "nucleus" S_1 also point outward. The bounding surface S_2 is a bit wierd: the surface that abuts the "nucleus" has *inward* normals; the surface on the outer edge has normal pointing outward as normal. The inward normal on the inner surface ensures that when we combine surfaces S_1 and S_2, the inner bits cancel and only the original surface S is left over.

Whereas the flux through S is hard to compute, the fluxes through S_1 and S_2 are easy:

$$\begin{aligned} \Phi_1 &= \oint_{S_1} \vec{E} \cdot d\vec{A} = 4\pi q \\ \Phi_2 &= \oint_{S_2} \vec{E} \cdot d\vec{A} = 0 \ . \end{aligned}$$

The first result holds because the surface S_1 is a sphere, and we can use the flux through a sphere result we already worked out. The second result holds because the number of field lines entering the surface S_2 on the inner edge is exactly equal to the number that leave S_2 on the outer edge. The total flux through the surface S is therefore $4\pi q$:

$$\Phi = \oint_S \vec{E} \cdot d\vec{A} = 4\pi q \ .$$

Just like for a sphere.

What if there's more than one charge inside S? We just invoke superposition, and compute the flux from each individual charge; then we add the result. It's pretty obvious what we get:

$$\Phi = \oint_S \vec{E} \cdot d\vec{A} = 4\pi \sum_{i=1}^{N} q_i \ .$$

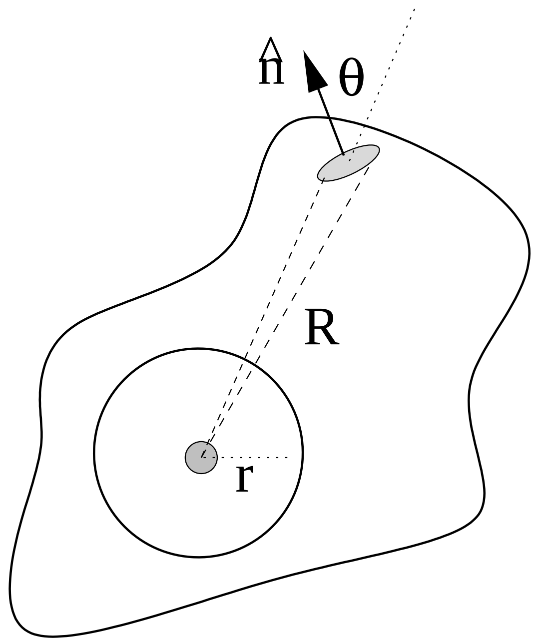
We have just worked out **Gauss's Law**: *The electric flux Φ through a closed surface S is equal to the total charge contained inside S.* Mathematically, we say

$$\Phi = \oint_S \vec{E} \cdot d\vec{A} = 4\pi q(\text{Contained within } S) \ .$$

This is in CGS units; in SI units, the factor 4π becomes $1/\epsilon_0$.

2.5 Gauss's law: More careful proof

Let's revisit the Gaussian amoeba and show the flux is $4\pi q$ independent of the surface's shape a little more carefully. Consider a cone of solid angle $d\Omega$ that goes out from the charge through both the spherical "nucleus" and out to the "real" surface:



The cone intersects the sphere at radius r; the surface of intersection is a circle whose area is $da = r^2\,d\Omega$, oriented parallel to $\hat{r}$. The electric flux through this patch is

$$d\Phi_{\text{sphere}} = \frac{q\hat{r}}{r^2} \cdot r^2\,d\Omega\,\hat{r} = q\,d\Omega\ .$$

The cone intersects the outer surface at a radius R. The surface of intersection is a very small ellipse whose area is $R^2\,d\Omega/\cos\theta$, oriented parallel to $\hat{n}$, which is inclined at an angle θ with respect to $\hat{r}$. The electric flux through this patch is

$$\begin{aligned} d\Phi_S &= \frac{q\hat{r}}{R^2} \cdot \frac{R^2\,d\Omega\,\hat{n}}{\cos\theta} \\ &= q\,d\Omega \frac{\hat{n}\cdot\hat{r}}{\cos\theta} \\ &= q\,d\Omega\ . \end{aligned}$$

The fluxes through each differential patch are equal; integrating over the total solid angle (4π steradians) subtended by each surface, we find the same flux: $\Phi_S = \Phi_{\text{sphere}} = 4\pi q$.

The generalization to multiple charges is then simple; Gauss's law follows.

2.6 Using Gauss's law

As I hope you'll come to appreciate, Gauss's law is our friend: in situations with lots of symmetry it makes calculating $\vec{E}$ really easy. The caveat is that it is *only* useful for this purpose when there's enough symmetry. Gauss's law is *always* true, but it is not always useful in this way.

The simplest example is the case of spherical symmetry.

2.6.1 Example: Spherical symmetry

Suppose we've got some perfectly spherical distribution of charge, with density $\rho(r)$. (Note that this density is a function of radius.) Enclose this distribution with a spherical surface S of radius r. Gauss's law tells us

$$\oint_S \vec{E} \cdot d\vec{A} = 4\pi q(\text{inside } S) \ .$$

Since the field is spherically symmetric, $\vec{E}$ must point in the radial direction: $\vec{E} \propto \hat{r}$. Further, it can only depend on r, which means that it is constant over any spherical surface. The integral is thus easy to evaluate:

$$\oint_S \vec{E} \cdot d\vec{A} = 4\pi r^2 E(r) \ .$$

The amount of charge contained inside S is given by integrating the charge density $\rho(r)$ over the volume V contained by S:

$$q = \int_V \rho(r)\, dV \ .$$

So,

$$\vec{E}(r) = \frac{\int_V \rho(r)\, dV}{r^2}\hat{r} = \frac{q(r)}{r^2}\hat{r} \ .$$

Let's specialize this further: put

$$\begin{aligned} \rho(r) &= \rho \qquad r \le a \\ &= 0 \qquad r > a \ . \end{aligned}$$

(ρ is a constant.) Then,

$$\begin{aligned} q(r) &= \int_V \rho(r)\, dV \\ &= \int_0^r 4\pi\rho\, r^2\, dr \\ &= \frac{4}{3}\pi r^3 \rho \qquad r \le a \\ &= \frac{4}{3}\pi a^3 \rho \qquad r > a \ . \end{aligned}$$

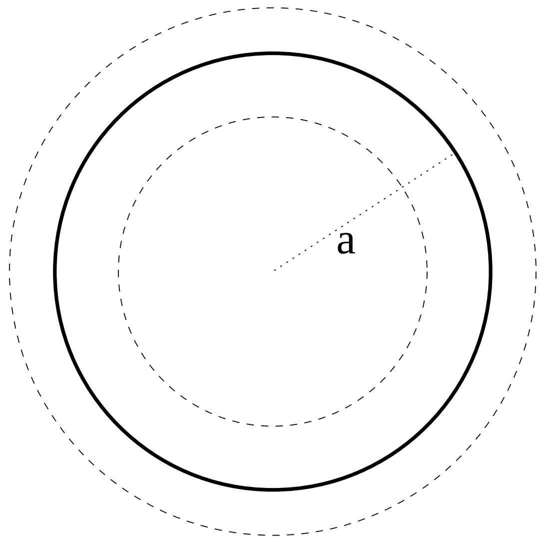
The electric field is

$$\begin{aligned} \vec{E}(r) &= \frac{4}{3}\pi\rho\, r\hat{r} \qquad r \le a \\ &= \frac{4\pi\rho a^3/3}{r^2}\hat{r} \qquad r > a \ . \end{aligned}$$

The electric field grows linearly with r inside the sphere, but falls off as $1/r^2$ outside the sphere. In fact, the external field is exactly that of a point charge $q = 4\pi\rho a^3/3$.

2.6.2 Example: Spherical symmetry 2

Suppose we've got a spherical shell of radius a, and a total charge q is distributed over the shell, so that the charge per unit area on the shell is $\sigma = q/4\pi a^2$. What is the electric field in this case?



Consider two Gaussian surfaces, one at $r < a$, one at $r > a$. For $r < a$, we have

$$\begin{aligned} 4\pi r^2 E(r) &= 4\pi q(r < a) = 0 \\ \rightarrow E(r) &= 0 \qquad r < a \ . \end{aligned}$$

There is no electric field *anywhere* inside the shell! On the outside, we have

$$\begin{aligned} 4\pi r^2 E(r) &= 4\pi q(r > a) = 4\pi q \\ \rightarrow E(r) &= \frac{q}{r^2} \qquad r > a \ . \end{aligned}$$

Outside, it looks just like a point charge sitting at the origin! The exterior of spherical charge distributions always look just like point charges at the origin.

Notice that, as we go from just barely inside the shell to just barely outside the shell, the electric field jumps suddenly:

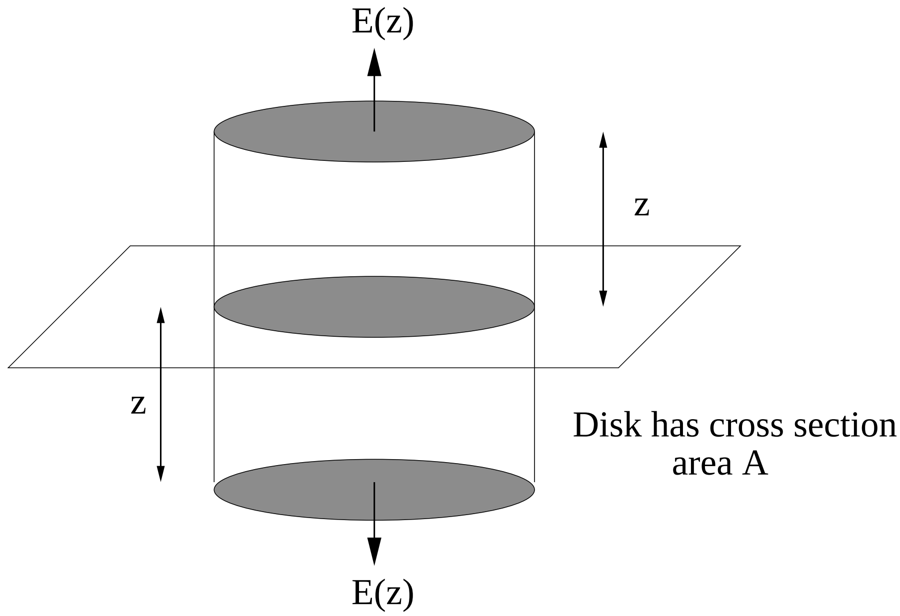
$$\Delta E = \frac{q}{a^2} - 0 = \frac{4\pi a^2 \sigma}{a^2} = 4\pi\sigma \ .$$

This discontinuous jump $\Delta E = 4\pi\sigma$ as we "cross" a sheet of charge per unit area σ is a general feature of such charge distributions. We see it in another example:

2.6.3 Example: Infinite sheet

Suppose we have a sheet of charge, lying in the (x, y) plane, and stretching out all the way to infinity. What is the electric field a distance z away from this sheet?

One way to do this would be to use the calculation of the disk we did earlier on, and take the limit $R \to \infty$. Here, we'll use Gauss's law: we make a "Gaussian pillbox" around the sheet like so:



Now, we think about the symmetry of the problem to figure out what Gauss's law tells us. The electric field cannot depend on our horizontal position — the sheet is the same in all directions, so all horizontal positions are identical. The field must also point either out (if it is positive charged) or in (negative). Finally, there can be no flux through the sides of the pillbox — the field can only point straight out by the planar symmetry. (If it did not, imagine rotating the plane — the field would change direction. That contradicts the notion that it is the same everywhere.) Gauss's law reduces to

$$\Phi = 2AE(z) .$$

The total charge contained inside the pillbox is σA, so we find

$$E(z) = 2\pi\sigma .$$

Notice three things here: first, this is the same field we found for the disk when z was very small. When we get extremely close to something like a disk, the infinite sheet calculation is *approximately* the electric field — essentially, you get so close that you can't "see" the edges of the disk, and it acts as though it were infinite. Second, this electric field is independent of distance — $\vec{E}$ is uniform all through space. Third, we see once again that, when we cross a sheet of uniform charge density, the *change* in the electric field is $4\pi\sigma$. In this case, it is $2\pi\sigma$ pointing down below, and $2\pi\sigma$ pointing up above.

Scott Hughes *8 February 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 3:
ELECTRIC FIELD ENERGY.
POTENTIAL; GRADIENT.
GAUSS'S LAW REVISITED; DIVERGENCE.

3.1 Energy in the field

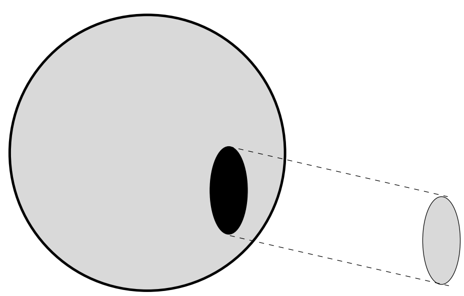
Suppose we take our spherical shell of radius r, with charge per unit area $\sigma = q/4\pi r^2$, and squeeze it. How much work does it take to do this squeezing? To answer this, we need to know how much pressure is being exerted by the shell's electric field on itself. Pressure is force/area; σ is charge/area; force is charge times electric field. Hence, our intuitive guess is that the pressure should be

$$\begin{aligned} P_{\rm guess} &= \text{(electric field)(charge/area)} \\ &= 4\pi\sigma^2 \ . \end{aligned}$$

(We have used the fact that the electric field just at the surface of the shell is $4\pi\sigma$.)

This is *almost* right, but contains a serious flaw. To understand the flaw, note that the electric field we used here, $E = 4\pi\sigma$, includes contributions from *all* of the charge on the sphere — including the charge that is being acted upon by the field. We must therefore be overcounting in some sense — a chunk of charge on the sphere cannot act upon itself.

It's not too hard to fix this up. Imagine pulling a little circular disk out of the sphere:



The *correct* formula for the pressure will be given by

$$\begin{aligned} P &= \text{(electric field in hole)(charge/area on disk)} \\ &= \text{(electric field in hole)}\sigma \ . \end{aligned}$$

We just need to figure out the electric field in the hole. Fortunately, there's a clever argument that spares us the pain of a hideous integral: we know that the field of the hole *plus* the field of the disk gives us the field of the sphere:

$$\begin{aligned} E_{\rm hole} + E_{\rm disk} &= E_{\rm sphere} \\ &= 4\pi\sigma \qquad \text{Just outside the sphere} \\ &= 0 \qquad \text{Just inside the sphere.} \end{aligned}$$

What's $E_{\rm disk}$? Fortunately, we only need to know this *very* close to the disk itself: from the calculations we did in Lecture 2, we have

$$\begin{aligned} E_{\rm disk} &= +2\pi\sigma \qquad \text{Outer side of disk} \\ &= -2\pi\sigma \qquad \text{Inner side of disk.} \end{aligned}$$

We're now done:

$$E_{\rm hole} = 2\pi\sigma \qquad \text{(Pointing radially outward)}$$

which means that the pressure is

$$P = 2\pi\sigma^2 \ .$$

(Purcell motivates the factor of 1/2 difference by pointing out that one gets the pressure by *averaging* the field at r over its discontinuity — the field is 0 at radius $r - \epsilon$, it is $4\pi\sigma$ at $r + \epsilon$. This argument is totally equivalent, and is in fact based upon Purcell problem 1.29.)

Now that we know the pressure it is easy to compute the work it takes to squeeze:

$$\begin{aligned} dW &= F\,dr \\ &= P\,4\pi r^2\,dr \\ &= 2\pi\sigma^2\,dV \ . \end{aligned}$$

On the last line, we've used the fact that when we compress the shell by dr, we have squeezed out a volume $dV = 4\pi r^2\,dr$. Now, using $E = 4\pi\sigma$, we can write this formula in terms of the electric field:

$$dW = \frac{E^2}{8\pi}\,dV \ .$$

The quantity $E^2/8\pi$ is the energy density of the electric field itself. (In SI units, we would have found $\epsilon_0 E^2/2$.) It shows up in this work calculation because when we squeeze the shell we are *creating* new electric field: the shell from $r - dr$ to r suddenly contains field, whereas it did not before. We had to put $dW = \text{(energy density)}\,dV$ amount of work into the system to make that field.

3.2 Electric potential difference

I move a charge q around in an electric field $\vec{E}$. The work done *by me* as I move this charge from position $\vec{a}$ to $\vec{b}$ is given by

$$\begin{aligned} U_{ba} &= \int_{\vec{a}}^{\vec{b}} \vec{F}_{\rm me}\cdot d\vec{s} \\ &= -\int_{\vec{a}}^{\vec{b}} \vec{F}_q\cdot d\vec{s} \\ &= -q\int_{\vec{a}}^{\vec{b}} \vec{E}\cdot d\vec{s}\ . \end{aligned}$$

$\vec{F}_{\rm me}$ is the force that I exert. By Newton's 3rd, this is equal and opposite to the force F_q that the electric field exerts on the charge. This is why we flip the sign going from line 1 to line

2. Make sure you understand this logic — that sign flip is an important part of stuff we're about to define! On the last line, we've expressed this in terms of the electric field. Recall that one of the things we learned about electric fields is that they generate *conservative* forces. This means that this integral must be independent of the path we take from $\vec{a}$ to $\vec{b}$. We'll return to this point later.

Since the work we do is proportional to the charge, we might as well divide it out. Doing so, we define the *electric potential difference* between $\vec{a}$ and $\vec{b}$:

$$\phi_{ba} = -\int_{\vec{a}}^{\vec{b}} \vec{E} \cdot d\vec{s} \,.$$

This function has a simple physical interpretation: it is the work that **I** must do, *per unit charge*, to move a charge from $\vec{a}$ to $\vec{b}$. Its units are energy/charge: erg/esu in cgs, Joule/Coulomb in SI. Both of these combinations are given special names: the cgs unit of potential is the *statvolt*, the SI unit is the *Volt*.

3.3 Electric potential

Let's say we fix the position $\vec{a}$ and consider the electric potential difference between $\vec{a}$ and a variety of different points $\vec{r}$. Since the point $\vec{a}$ is fixed, we can speak of the *electric potential* $\phi(\vec{r})$ relative to the *reference point* $\vec{a}$. The language we use here is very important — "potential" (as opposed to "potential difference") only makes sense provided we know what is the reference point. All potentials are then defined *relative to* that reference point.

In a large number of cases, it is very convenient to take the reference point $\vec{a}$ to be at infinity. (We'll discuss momentarily when it is not convenient to do so.) In this case, the electric potential at $\vec{r}$ is

$$\phi(\vec{r}) = -\int_{\infty}^{\vec{r}} \vec{E} \cdot d\vec{s} \,.$$

Suppose $\vec{E}$ arises from a point charge q at the origin:

$$\begin{aligned} \phi(\vec{r}) &= -\int_{\infty}^{\vec{r}} \vec{E} \cdot d\vec{s} \\ &= -\int_{\infty}^{r} \frac{q}{r^2} dr \\ &= \frac{q}{r} \,. \end{aligned}$$

Note that using this formula the potential difference has a very simple form:

$$\begin{aligned} \phi_{ba} &= -\int_{\vec{a}}^{\vec{b}} \vec{E} \cdot d\vec{s} \,. \\ &= \frac{q}{b} - \frac{q}{a} \\ &= \phi(\vec{b}) - \phi(\vec{a}) \,. \end{aligned}$$

As the name suggests, the potential difference is indeed just the difference in the potential between the two points.

By invoking superposition, we can generalize these formulas very simply: Suppose we have N point charges. Using infinity as our reference point, the potential at some point P is given by

$$\phi(P) = \sum_{i=1}^{N} \frac{q_i}{r_i} \, ,$$

where r_i is the distance from P to the ith charge. We can further generalize this to continuous distributions: for a charge density ρ distributed through a volume V,

$$\phi(P) \;=\; \int_V \frac{\rho \, dV}{r} \, ,$$

where r is the distance from a volume element dV to the field point; for surface charge density σ distributed over an area A,

$$\phi(P) \;=\; \int_A \frac{\sigma \, dA}{r} \, ;$$

for a line charge density λ distributed over a length L,

$$\phi(P) \;=\; \int_L \frac{\lambda \, dl}{r} \, .$$

3.4 An important caveat

There is one *massive* caveat that must be borne in mind when using these formulas: **they ONLY work if it is OK to set the reference point to infinity**. That, in turn only works if the charge distribution *itself* is of finite size. As a counterexample, consider an infinite line charge, with charge per unit length λ. Using Gauss's law, it is simple to show that the electric field of this charge distribution is given by

$$\vec{E} = \frac{2\lambda}{r}\hat{r} \, ,$$

where $\hat{r}$ points out from the line charge, and r is the distance to the line. Let's first work out the potential difference between two points r_1 and r_2 away from the line:

$$\begin{aligned}
\phi_{21} &= -\int_{r_1}^{r_2} \vec{E} \cdot d\vec{s} \\
&= -2\lambda \int_{r_1}^{r_2} \frac{dr}{r} \\
&= 2\lambda \ln\left(r_1/r_2\right) \, .
\end{aligned}$$

Suppose we make r_1 our reference point. Now, look what happens as we let this reference point approach infinity — the potential diverges! For a charge distribution like this, the field does not fall off quickly enough as a function of distance for infinity to be a useful reference point. In cases like this, we need to pick some other reference point. In such a situation, it is crucial that you make clear what you have picked.

3.5 Energy of a charge distribution

If we can use infinity as a reference point, then we can write down a formula for the energy of a charge distribution that is sometimes quite handy. From an earlier lecture, we worked out the formula

$$U = \frac{1}{2}\sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \ .$$

A useful way to rearrange this formula is as follows:

$$\begin{aligned} U &= \frac{1}{2}\sum_i q_i \sum_{j \neq i} \frac{q_i}{r_{ij}} \\ &= \frac{1}{2}\sum_i q_i \phi(r_i) \ . \end{aligned}$$

On the second line, we have used the fact that the quantity under the second sum is just the electric potential of all the charges *except* q_i, evaluated at q_i's location.

Now, we take a continuum limit: we put $q_i \to \rho\, dV$ and let the sum become an integral. We find

$$U = \frac{1}{2}\int \rho\phi\, dV \ .$$

This formula is *entirely* equivalent to other formulas we worked out for the energy of a charge distribution. Bear in mind, though, that we have implicitly assumed that the potential goes to zero infinitely far from the charge distribution!

3.6 Field from the potential

Consider the potential difference between the point $\vec{r}$ and $\vec{r} + \Delta\vec{r}$:

$$\begin{aligned} \Delta\phi &= -\int_{\vec{r}}^{\vec{r}+\Delta\vec{r}} \vec{E} \cdot d\vec{s} \\ &\simeq -\vec{E}(\vec{r}) \cdot \Delta\vec{r} \ . \end{aligned}$$

Now, expand the vectors in Cartesian coordinates:

$$\begin{aligned} \Delta\phi &\simeq -(E_x\hat{x} + E_y\hat{y} + E_z\hat{z}) \cdot (\Delta x\hat{x} + \Delta y\hat{y} + \Delta z\hat{z}) \\ &\simeq -E_x\,\Delta x - E_y\,\Delta y - E_z\,\Delta z \ . \end{aligned}$$

Dividing, taking the limit $\Delta \to 0$, and invoking the definition of the partial derivative, we find

$$E_x = -\frac{\partial\phi}{\partial x} \ , \qquad E_y = -\frac{\partial\phi}{\partial y} \ , \qquad E_z = -\frac{\partial\phi}{\partial z} \ ;$$

or

$$\begin{aligned} \vec{E} &= -\hat{x}\frac{\partial\phi}{\partial x} - \hat{y}\frac{\partial\phi}{\partial y} - \hat{z}\frac{\partial\phi}{\partial z} \\ &\equiv -\vec{\nabla}\phi \ . \end{aligned}$$

On the last line we have defined the *gradient operator*,

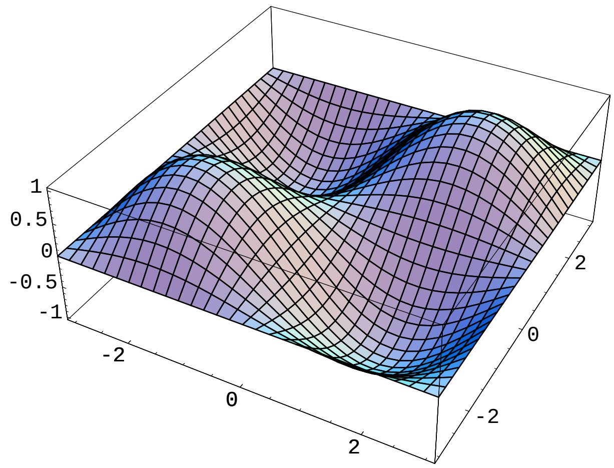
$$\vec{\nabla} = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z} \ .$$

The electric field is thus given by the *gradient* of the potential. You will sometimes see this written $\vec{E} = -\mathbf{grad}\,\phi$.

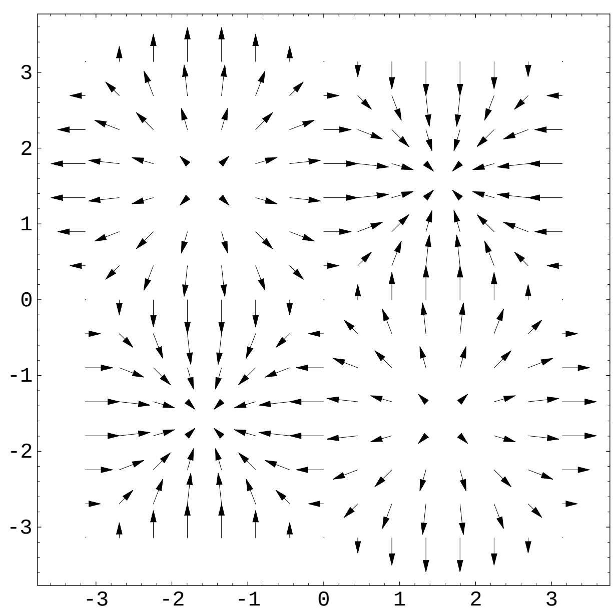
3.7 The gradient and equipotentials

We've defined a mathematical operation which we named the gradient. What exactly *is* it? To get some intuition, let's imagine a one dimensional potential, $\phi(x)$. The gradient is given by $\vec{\nabla}\phi = \hat{x}\, d\phi/dx$. Bearing in mind that the derivative is the function's *slope*, the gradient appears to tell us the direction in which the function changes, and the rate of that change.

A 1-D example is simple, but also a touch silly. Let's try a 2-D example: put $\phi(x, y) = \sin(x)\sin(y)$. This potential looks like

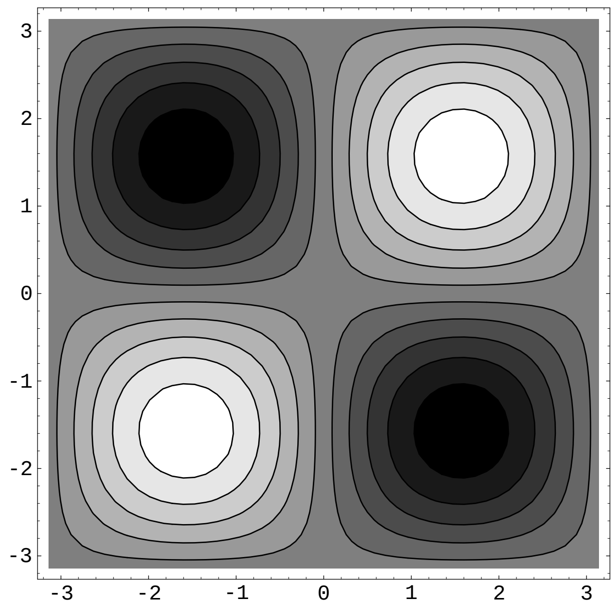


It is simple to calculate this potential's gradient: $\vec{\nabla}\phi = \hat{x}\,\cos(x)\sin(y) + \hat{y}\,\sin(x)\cos(y)$. This vector field looks like



Notice the way the little arrows are pointing. Our intuition from the silly 1-D example is exactly right: the gradient points in the direction in which the function increases fastest. At the minima of the function (upper left, lower right), the gradient points out. At the maxima (upper right, lower left), the gradient points in. A simple way to summarize this is to remember that the gradient always points "uphill". Bearing in mind that the electric field is the *negative* gradient of ϕ, we can say that the electric field points "downhill".

Another useful concept we can build from these plots is the notion of an *equipotential surface*. As the name suggests, an equipotential surface is just a surface along which the potential takes some constant value. For our 2-D example, the equipotential "surfaces" are really curves in the $x-y$ plane:



We can deduce one further important property of equipotential surfaces. By definition, the potential is constant on them, so they define the directions along which the potential does not change at all. The gradient, by contrast, defines the direction along which the potential changes *the most*! If you think about this for a moment, it becomes clear the gradient must be exactly perpendicular to equipotential surfaces. Since the electric field is $-\vec{\nabla}\phi$, it follows that the electric field is also exactly perpendicular to equipotentials. A very handy way of putting this is to say that electric field lines are orthogonal to equipotential surfaces.

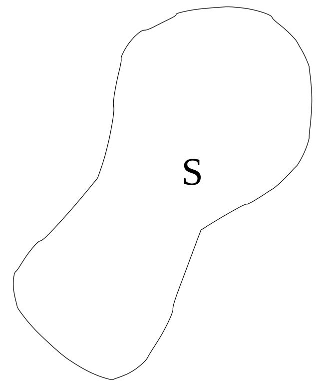
(If you are a hiker and you have ever used a topographic map, much of this should seem quite familiar. The contours of a topo map label surfaces of constant height, which is exactly related to the *gravitational* potential. If you move perpendicular to the contour lines, you are moving along the steepest slope possible — desperately scrambling up, or rather quickly tumbling down.)

3.8 Divergence

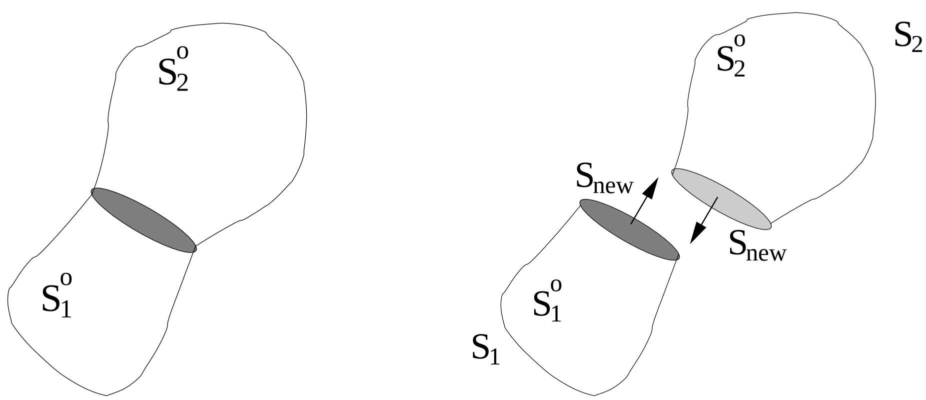
The gradient which we just learned about is the first and simplest mathematical operation in *vector calculus*. We're going to continue this theme for a little while, exploring several vector calculus operations which we will use exhaustively in this course.

To set up the next vector calculus operation, let's return to the notion of flux. Suppose we've got some vector field $\vec{F}$, and we want to calculate its flux Φ through a closed surface S:

$$\Phi = \oint_S \vec{F} \cdot d\vec{A} \, .$$



Now, we do something that may seem kind of silly: we cut the volume into two pieces.



(For clarity, the two halves on the right are drawn separated. We really leave the two pieces right next to each other.) Before calculating, let's make sure we understand what the various surfaces in this diagram mean. On the left, we have the surfaces $S_1^{\rm o}$ and $S_2^{\rm o}$. These are just the top-right and bottom-left halves of the "old" surface S. Note that these are *not closed* surfaces. On the right, we still have $S_1^{\rm o}$ and $S_2^{\rm o}$; in addition, we have a pair of surfaces labeled $S_{\rm new}$. These surfaces are exactly the same area but are oriented in opposite directions. S_1 and S_2 are the *closed* surfaces surrounding the two halves of the blob.

OK, so what's the flux in this case? It must be exactly the same, Φ, since everything's still the same as far as the field $\vec{F}$ is concerned. But the way we calculate it is different:

$$\begin{aligned}
\Phi &= \oint_S \vec{F}\cdot d\vec{A} \\
&= \int_{S_1^{\rm o}} \vec{F}\cdot d\vec{A} + \int_{S_2^{\rm o}} \vec{F}\cdot d\vec{A} \\
&= \int_{S_1^{\rm o}} \vec{F}\cdot d\vec{A} + \int_{S_2^{\rm o}} \vec{F}\cdot d\vec{A} + \int_{S_{\rm new}} \vec{F}\cdot d\vec{A} - \int_{S_{\rm new}} \vec{F}\cdot d\vec{A} \\
&= \left(\int_{S_1^{\rm o}} \vec{F}\cdot d\vec{A} + \int_{S_{\rm new}} \vec{F}\cdot d\vec{A}\right) + \left(\int_{S_2^{\rm o}} \vec{F}\cdot d\vec{A} + \int_{S_{\rm new}} \vec{F}\cdot (-d\vec{A})\right) \\
&= \oint_{S_1} \vec{F}\cdot d\vec{A} + \oint_{S_2} \vec{F}\cdot d\vec{A} \\
&\equiv \Phi_1 + \Phi_2\ .
\end{aligned}$$

Let's go through this carefully. On the second line, we have just broken the closed surface S into the two open surfaces S_1^o and S_2^o. On the third line, we have added and subtracted an integral over S_{new}. On the fourth line, we've gathered these terms together in a special way. This special way shows that the combined surfaces are just the *closed* surfaces S_1 and S_2. On the sixth line, we define these two integrals as Φ_1 and Φ_2.

We can keep dividing this volume up into smaller and smaller subvolumes; everytime we do so, there will be facing surfaces whose orientations will cancel each other out. This allows us to break the total flux up into contributions that come from the surfaces of all these little subvolumes:

$$\Phi = \sum_{i=1}^{\text{big}} \Phi_k \ .$$

Since we can keep chopping it up into little volumes, the total flux should be proportional to the volume. This motivates the definition of the *divergence*:

$$\mathbf{div}\,\vec{F} = \lim_{V\to 0} \frac{\oint_S \vec{F}\cdot d\vec{A}}{V} \ .$$

In words, the divergence is what you get when you compute the flux through the surface S bounding a volume V, *divided by* V, in the limit in which V goes to zero.

3.9 Gauss's theorem and Gauss's law

Now that we have defined the divergence, we are set to prove a very important theorem. We start with this equation again,

$$\Phi = \sum_{i=1}^{\text{big}} \Phi_k \ ,$$

but now we write out the integrals:

$$\begin{aligned}
\int_S \vec{F}\cdot d\vec{A} &= \sum_{i=1}^{N} \oint_{S_i} \vec{F}\cdot d\vec{A}_i \\
&= \sum_{i=1}^{N} V_i \frac{\oint_{S_i} \vec{F}\cdot d\vec{A}_i}{V_i} \\
&= \sum_{i=1}^{N} V_i\,\mathbf{div}\,\vec{F} \\
&= \int_V \mathbf{div}\,\vec{F}\,dV \ .
\end{aligned}$$

In words, the flux of a vector function $\vec{F}$ through a closed surface S equals the integral of $\mathbf{div}\,\vec{F}$ over the volume enclosed by S. This is called *Gauss's theorem.*

As a concrete example, let's take $\vec{F}$ to be the electric field $\vec{E}$. We thus have

$$\oint_S \vec{E}\cdot d\vec{A} = \int_V \mathbf{div}\,\vec{E}\,dV \ .$$

However, we proved a while ago that

$$\begin{aligned}\oint_S \vec{E}\cdot d\vec{A} &= 4\pi Q(\text{inside } S)\\ &= 4\pi\int_V \rho\, dV\ .\end{aligned}$$

Combining these, we have

$$\int_V \mathbf{div}\,\vec{E}\, dV = 4\pi\int_V \rho\, dV\ ,$$

or

$$\int_V \left(\mathbf{div}\,\vec{E} - 4\pi\rho\right) dV = 0\ .$$

This equation is only true for all volumes V only if

$$\boxed{\boxed{\mathbf{div}\,\vec{E} = 4\pi\rho}}$$

This equation also expresses Gauss's law, only in differential (rather than integral) form. Learn it well! In SI units, in keeping with the rule that the electric field carries a factor $1/4\pi\epsilon_0$, this formula becomes

$$\mathbf{div}\,\vec{E} = \rho/\epsilon_0 \qquad \text{(SI)}\ .$$

Scott Hughes *10 February 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 4:
WRAP UP OF VECTOR CALCULUS:
POISSON & LAPLACE EQUATIONS; CURL

4.1 Summary: Vector calculus so far

We have learned several mathematical operations which fall into the category of vector calculus. In Cartesian coordinates, these operations can be written in very compact form using the following operator:

$$\vec{\nabla} \equiv \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z} \; .$$

The first vector calculus operation we learned is the *gradient*. It converts the electric *potential* into the electric *field*:

$$\begin{aligned}\vec{E} &= -\mathbf{grad}\,\phi \\ &= -\vec{\nabla}\phi \; .\end{aligned}$$

The second operation is the *divergence*, which relates the electric field to the charge density:

$$\mathbf{div}\,\vec{E} = 4\pi\rho \; .$$

Via *Gauss's theorem* (also known as the *divergence theorem*), we can relate the flux of any vector field $\vec{F}$ through a closed surface S to the integral of the divergence of $\vec{F}$ over the volume enclosed by S:

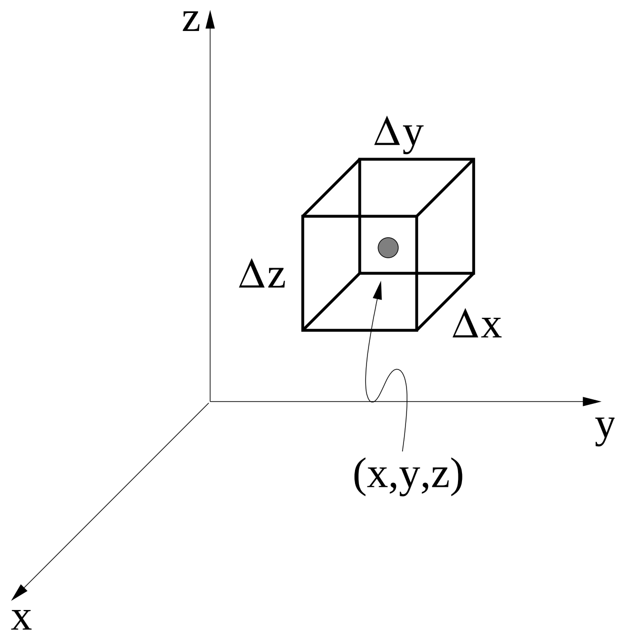
$$\oint_S \vec{F}\cdot d\vec{A} = \int_V \mathbf{div}\,\vec{F}\,dV \; .$$

This follows from the definition of the divergence:

$$\mathbf{div}\,\vec{F} = \lim_{V\to 0}\frac{\oint_S \vec{F}\cdot d\vec{A}}{V} \; .$$

4.2 Divergence in Cartesian coordinates

So far, we've only defined the divergence as a particular limit. We now want to develop a concrete calculation showing its value. To do so, consider an infinitesimal cube with sides Δx, Δy, and Δz, centered on the coordinate x, y, z:



[The large dot in the center of this cube is only there to mark the coordinate (x, y, z) — it doesn't mean anything else.] We take this region to be filled with a vector field $\vec{F}$. To begin, we want to compute the flux of $\vec{F}$ over this cube. Let's look at the flux over just the z faces (the top and bottom):

$$\begin{aligned}\Delta\Phi_z &= \int_{\text{top \& bottom}} \vec{F}\cdot d\vec{A} \\ &\simeq \Delta x\Delta y\left[F_z\left(x,y,z+\Delta z/2\right)-F_z\left(x,y,z-\Delta z/2\right)\right]\ .\end{aligned}$$

Since the cube is taken to be very small (indeed, we will soon take the limit $\Delta x, \Delta y, \Delta z \to 0$), we approximate the integral. Notice that the contribution from the bottom face $(z - \Delta z/2)$ enters with a minus sign, consistent with that face pointing down.

Now, to evaluate the function under the integrand, we make a Taylor expansion about z:

$$\begin{aligned}\Delta\Phi_z &\simeq \Delta x\Delta y\left[F_z\left(x,y,z\right)+\frac{\Delta z}{2}\frac{\partial F_z}{\partial z}\right]-\Delta x\Delta y\left[F_z\left(x,y,z\right)-\frac{\Delta z}{2}\frac{\partial F_z}{\partial z}\right] \\ &\simeq \Delta x\Delta y\Delta z\frac{\partial F_z}{\partial z}\ .\end{aligned}$$

Repeating this exercise to find the fluxes through the other four sides, we find

$$\begin{aligned}\Delta\Phi_x &\simeq \Delta x\Delta y\Delta z\frac{\partial F_x}{\partial x} \\ \Delta\Phi_y &\simeq \Delta x\Delta y\Delta z\frac{\partial F_y}{\partial y}\ .\end{aligned}$$

We now have enough information to evaluate the divergence of $\vec{F}$:

$$
\begin{aligned}
\mathbf{div}\,\vec{F} &= \lim_{V\to 0} \frac{\int \vec{F}\cdot d\vec{A}}{V} \\
&= \lim_{\Delta x,\Delta y,\Delta z\to 0} \frac{\Delta x\Delta y\Delta z\left(\partial F_x/\partial x + \partial F_y/\partial y + \partial F_z/\partial z\right)}{\Delta x\Delta y\Delta z} \\
&= \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \\
&\equiv \vec{\nabla}\cdot\vec{F}\ .
\end{aligned}
$$

On the last line, we've taken advantage of the fact that we can think of the gradient operator as a vector. In Cartesian coordinates, the divergence is nothing more than combination of derivatives one finds by taking the dot product of this operator with a vector field. It should be *strongly* emphasized at this point, however, that this only works in Cartesian coordinates. In spherical coordinates or cylindrical coordinates, the divergence is not just given by a dot product like this!

4.2.1 Example: Recovering ρ from the field

In Lecture 2, we worked out the electric field associated with a sphere of radius a containing uniform charge density ρ:

$$
\begin{aligned}
\vec{E}(r) &= \frac{4}{3}\pi\rho r\hat{r} \qquad r \le a \\
&= \frac{4\pi\rho a^3/3}{r^2}\hat{r} \qquad r > a\ .
\end{aligned}
$$

Recall that we found this by applying the integral form of Gauss's law. Breaking it up into Cartesian components, we have for $r \le a$

$$
\begin{aligned}
E_x &= \frac{4}{3}\pi\rho x \\
E_y &= \frac{4}{3}\pi\rho y \\
E_z &= \frac{4}{3}\pi\rho z\ ;
\end{aligned}
$$

for $r > a$,

$$
\begin{aligned}
E_x &= \frac{4\pi\rho a^3 x/3}{\left(x^2+y^2+z^2\right)^{3/2}} \\
E_y &= \frac{4\pi\rho a^3 y/3}{\left(x^2+y^2+z^2\right)^{3/2}} \\
E_z &= \frac{4\pi\rho a^3 z/3}{\left(x^2+y^2+z^2\right)^{3/2}}\ .
\end{aligned}
$$

(We used the facts that $r\hat{r} = x\hat{x} + y\hat{x} + z\hat{z}$, and $\hat{r}/r^2 = r\hat{r}/r^3$ in here.)

OK, let's now evaluate $\vec{\nabla}\cdot\vec{E}$. For $r \le a$, we have

$$
\frac{\partial E_x}{\partial x} = \frac{\partial E_y}{\partial y} = \frac{\partial E_z}{\partial z} = \frac{4}{3}\pi\rho\ ,
$$

so

$$\vec{\nabla} \cdot \vec{E} = \frac{4}{3}\pi\rho + \frac{4}{3}\pi\rho + \frac{4}{3}\pi\rho = 4\pi\rho$$

exactly as advertised! The region $r > a$ is a bit messier:

$$\begin{aligned} \frac{\partial E_x}{\partial x} &= \frac{\partial}{\partial x}\frac{4\pi\rho a^3 x/3}{(x^2+y^2+z^2)^{3/2}} && (1) \\ &= \frac{4\pi\rho a^3}{3}\left[\frac{1}{(x^2+y^2+z^2)^{3/2}} - \frac{3x^2}{(x^2+y^2+z^2)^{5/2}}\right] . && (2) \end{aligned}$$

Repeating for y and z components and adding, we have

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{4\pi\rho a^3}{3}\left[\frac{3}{(x^2+y^2+z^2)^{3/2}} - \frac{3x^2+3y^2+3z^2}{(x^2+y^2+z^2)^{5/2}}\right] && (3) \\ &= \frac{4\pi\rho a^3}{3}\left[\frac{3}{(x^2+y^2+z^2)^{3/2}} - \frac{3}{(x^2+y^2+z^2)^{3/2}}\right] && (4) \\ &= 0 . && (5) \end{aligned}$$

Since there is no charge in the exterior region — $\rho = 0$ for $r > a$ — this is exactly what $\vec{\nabla} \cdot \vec{E}$ *should* give.

4.3 Laplacian operator and Poisson's equation

The div and grad operations can be combined to make an equation that relates the potential ϕ to the charge density ρ:

$$\textbf{div}\,\textbf{grad}\,\phi = -4\pi\rho .$$

In Cartesian coordinates, this takes a very nice, simple form:

$$\vec{\nabla} \cdot (\vec{\nabla}\phi) = -4\pi\rho$$

or

$$\nabla^2\phi = -4\pi\rho .$$

The operator ∇^2 is, in Cartesian coordinates, exactly what'd you expect taking $\vec{\nabla}$ and dotting it into itself:

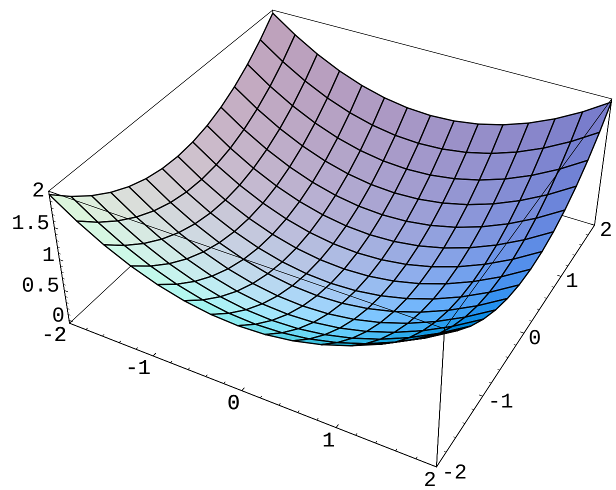
$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} .$$

The equation $\nabla^2\phi = -4\pi\rho$ is called Poisson's equation; the operator ∇^2 is called the "Laplacian".

For a two dimensional function $\phi(x, y)$, the Laplacian tells us about the *average* curvature of the function. Let's look at a couple of examples. First, take a potential that is a bowl-like paraboloid in the $x - y$ plane:

$$\phi(x, y) = \frac{a}{4}\left(x^2 + y^2\right) \ ,$$

which has the appearance



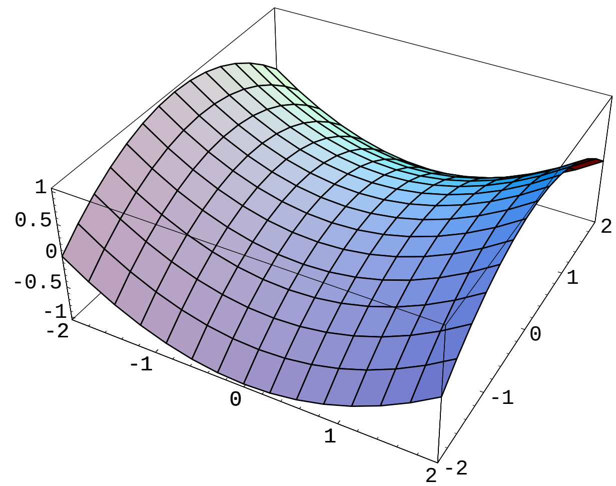
For this function, $\nabla^2\phi = \partial^2\phi/\partial x^2 + \partial^2\phi/\partial y^2 = a$ — the function has positive curvature everywhere, reflected by the fact that the bowl-like function curls up everywhere. There is a minimum at $x = 0$, $y = 0$.

4.4 Laplace's equation

Second, look at a potential that has a hyperboloidal shape in the plane:

$$\phi(x, y) = \frac{a}{4}\left(x^2 - y^2\right) \ .$$

This function has the appearance



In this case, we find $\nabla^2\phi = \partial^2\phi/\partial x^2 + \partial^2\phi/\partial y^2 = 0$. The average curvature of this function is zero — it curves up in x, and down in y. There is *no* minimum (or maximum) in this case, only a saddle point at $x = 0$, $y = 0$.

The case $\nabla^2\phi = 0$ — that is, the charge density $\rho = 0$ — is a special case of Poisson's equation called *Laplace's* equation. Laplace didn't just attach his name to Poisson's equation in this special case out of some kind of greedy spite. The two equations really merit separate treatment and different names, since their solutions (and methods for solving them) have rather different properties. (Indeed, Laplace was Poisson's teacher; the Laplace equation predates Poisson's by several decades.)

4.5 Earnshaw's Theorem

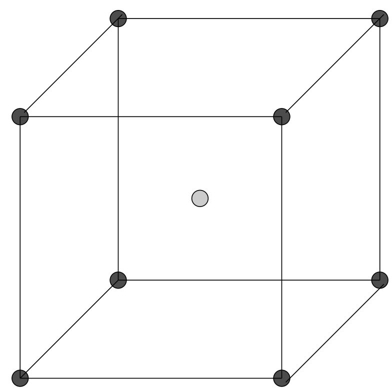
A potential that satisfies Laplace's equation has a very important property: if $\nabla^2\phi = 0$ in some region, then ϕ has no local minimum (or maximum) in that region. For there to be a minimum, a function must have positive second derivative in every possible direction. This just isn't possible if

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0 \ .$$

Suppose we manage to locate a place where ϕ has a minimum in the x direction and a minimum in the y direction. We know, then, that $\partial^2\phi/\partial x^2 > 0$ and $\partial^2\phi/\partial y^2 > 0$. The only way for ϕ to satisfy Laplace's equation is if $\partial^2\phi/\partial z^2 < 0$. This location is a minimum in x and y, but it is *not* a minimum in z. The best we can do is find a saddle point, a location that is a minimum along two directions, but a maximum along the third.

The non-existence of minima for regions in which $\rho = 0$ has an important physical consequence: it means that it is impossible to hold a charge in stable equilibrium with electrostatic fields. Any object that is in stable equilibrium must be at a minimum of its potential energy. Since the potential energy of a charge in an electrostatic field is just $U = q\phi$, it follows that stable equilibrium requires a minimum in the electric potential. Such a minimum just isn't possible when $\rho = 0$ — we *must* have $\nabla^2\phi = 0$. This result goes under the name *Earnshaw's Theorem*.

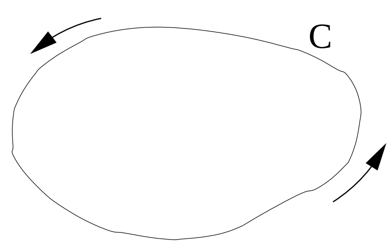
This means that any configuration make out of pure point charges cannot be stable. For example, if we glue positive charges to the corners of a cube, it is impossible to stably hold any kind of charge inside that cube:



All the forces cancel out on a charge at the exact center — but, if we displace it just slightly away from the center, it will rapidly fall out of equilibrium and come shooting out of the cube. This result was very disturbing to scientists in the late 19th and early 20th centuries, since it implied that atoms could not be stable — the electron should fall into the proton after a few microseconds. It wasn't until we developed an understanding of quantum mechanics that the stability of matter made any sense.

4.6 The curl

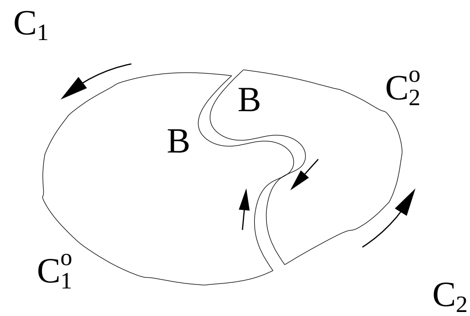
We wrap up this lecture with the final instrument in our vector calculus arsenal. We begin by thinking about the line integral of a vector field $\vec{F}$ around some closed loop C:



This quantity is called the *circulation* Γ of $\vec{F}$:

$$\Gamma = \oint_C \vec{F} \cdot d\vec{s}\,.$$

We now bridge the contour C with a new path B. Doing so we make two new loops C_1 and C_2. C_1 is built from an open loop C_1^o and the path B going upwards; C_2 is built from an open loop C_2^o and the path B going downwards:



Now, we manipulate the integral in a manner similar to what we did when we derived the divergence:

$$\begin{aligned}\Gamma &= \int_{C_1^o} \vec{F}\cdot d\vec{s} + \int_{C_2^o} \vec{F}\cdot d\vec{s} \\ &= \int_{C_1^o} \vec{F}\cdot d\vec{s} + \int_{C_2^o} \vec{F}\cdot d\vec{s} + \int_B \vec{F}\cdot d\vec{s} + \int_B \vec{F}\cdot(-d\vec{s}) \\ &= \left(\int_{C_1^o} \vec{F}\cdot d\vec{s} + \int_B \vec{F}\cdot d\vec{s}\right) + \left(\int_{C_2^o} \vec{F}\cdot d\vec{s} - \int_B \vec{F}\cdot d\vec{s}\right) \\ &= \oint_{C_1} \vec{F}\cdot d\vec{s} + \oint_{C_2} \vec{F}\cdot d\vec{s} \\ &\equiv \Gamma_1 + \Gamma_2\ .\end{aligned}$$

The key bit to this analysis is that when we cut a big contour in half, we end up with a path going in one direction on the new "bridging" contour, and in the opposite direction on the other side. Hence, everytime we do this, we get "internal" contours that cancel out, and what's left is equivalent to the original big contour.

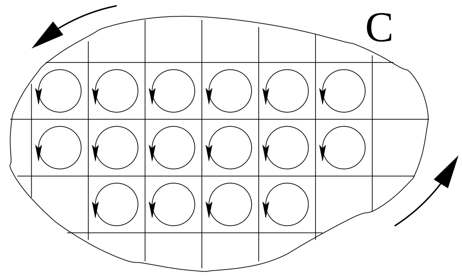
We now repeatedly break up the contour into lots of little bits:

$$\oint_C \vec{F}\cdot d\vec{s} = \sum_{i=1}^{\rm big} \oint_{C_i} \vec{F}\cdot d\vec{s}\ ,$$

or

$$\Gamma = \sum_{i=1}^{\rm big} \Gamma_i\ .$$

As we keep breaking the object up, we end up covering the *area* enclosed by the contour with a grid of little circulation cells:



and so the total circulation must be proportional to the area that the contour encloses. We thus define a vector called the *curl* of $\vec{F}$:

$$\mathbf{curl}\,\vec{F}\cdot\hat{n} = \lim_{A\to 0}\frac{\oint_C \vec{F}\cdot d\vec{s}}{A}\ .$$

The area A is bounded by the contour C. The curl is thus circulation of $\vec{F}$ per unit A, in the limit in which A goes to zero. Notice that the curl is a *vector*: it points in the direction $\hat{n}$ which is normal to the surface. We define this direction with the right hand rule: curl your right-hand fingers in the direction of C's circulation, and your thumb defines $\hat{n}$.

As with the divergence, an important theorem is connected to the curl. We start with its definition, and then do the breaking up into little bits:

$$\begin{aligned}
\Gamma = \oint_C \vec{F}\cdot d\vec{s} &= \sum_{i=1}^{\text{big}}\oint_{C_i}\vec{F}\cdot d\vec{s}\\
&= \sum_{i=1}^{\text{big}} A_i\frac{\oint_{C_i}\vec{F}\cdot d\vec{s}}{A_i}\\
&= \sum_{i=1}^{\text{big}} A_i\mathbf{curl}\,\vec{F}\cdot\hat{n}\\
&= \sum_{i=1}^{\text{big}} \mathbf{curl}\,\vec{F}\cdot(A_i\,\hat{n})\\
&= \oint_A \mathbf{curl}\,\vec{F}\cdot d\vec{A}\ .
\end{aligned}$$

In words, the circulation of a vector field $\vec{F}$ over a closed contour C equals the flux of the curl of $\vec{F}$ through the surface bounded by C. This rule is known as *Stoke's theorem.*

Suppose our vector field $\vec{F}$ is just an electrostatic field, $\vec{E}$. We already know that

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

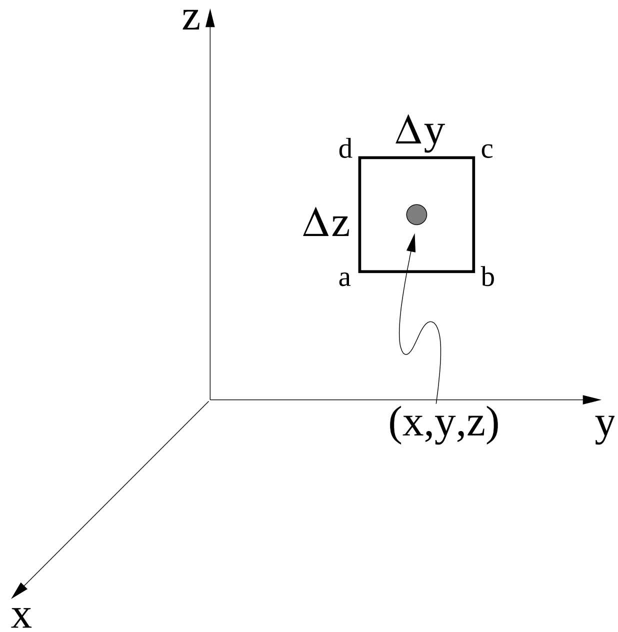
for any contour C since the electric force is conservative. It follows from this that

$$\boxed{\mathbf{curl}\,\vec{E} = 0}$$

This is an important point, which is worth emphasizing: a "good" electrostatic field has *no curl.*

4.7 The curl in Cartesian coordinates

Consider a square lying in a plane parallel to the $y-z$ plane, with sides Δy and Δz, centered on the coordinate (x, y, z). This square is immersed in a vector field $\vec{F}$. Using this square, let's work out what the curl must be.



We approximate the integral along the 4 legs as follows:

$$\begin{aligned}
\int_a^b \vec{F}\cdot d\vec{s} &= F_y(x,y,z-\Delta z/2)\Delta y \simeq \left[F_y(x,y,z) - \frac{\Delta z}{2}\frac{\partial F_y}{\partial z}\right]\Delta y \\
\int_b^c \vec{F}\cdot d\vec{s} &= F_z(x,y+\Delta y/2,z)\Delta z \simeq \left[F_z(x,y,z) + \frac{\Delta y}{2}\frac{\partial F_z}{\partial y}\right]\Delta z \\
\int_c^d \vec{F}\cdot d\vec{s} &= F_y(x,y,z+\Delta z/2)(-\Delta y) \simeq -\left[F_y(x,y,z) + \frac{\Delta z}{2}\frac{\partial F_y}{\partial z}\right]\Delta y \\
\int_d^a \vec{F}\cdot d\vec{s} &= F_z(x,y-\Delta y/2,z)(-\Delta z) \simeq -\left[F_z(x,y,z) - \frac{\Delta y}{2}\frac{\partial F_z}{\partial y}\right]\Delta z \;.
\end{aligned}$$

Adding these together, we have

$$\oint_{abcd} \vec{F}\cdot d\vec{s} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\Delta y \Delta z \;,$$

which tells us that

$$\begin{aligned}
\left[\mathbf{curl}\,\vec{F}\right]_x &= \lim_{\Delta z,\Delta y\to 0} \frac{\oint_{abcd}\vec{F}\cdot d\vec{s}}{\Delta z \Delta y} \\
&= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) .
\end{aligned}$$

This is the x component of the curl because, by right-hand rule, that is the orientation of the square.

If we repeat this exercise for squares oriented parallel to the $x - y$ planes and the $x - z$ planes, we find

$$\begin{aligned}
\left[\mathbf{curl}\,\vec{F}\right]_y &= \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) \\
\left[\mathbf{curl}\,\vec{F}\right]_z &= \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) .
\end{aligned}$$

There's a nice trick that we can use to work out these two components: we use *cyclic permutation* to flip around the axes. This means we just cyclic the labels, which is equivalent to rotating the axes around. To get $\left[\mathbf{curl}\,\vec{F}\right]_y$, take the answer for $\left[\mathbf{curl}\,\vec{F}\right]_x$, and permute the coordinate indices: x becomes y, y becomes z, z becomes x. Do it one more time, and you've got the answer for $\left[\mathbf{curl}\,\vec{F}\right]_z$.

Putting all of this together, we have

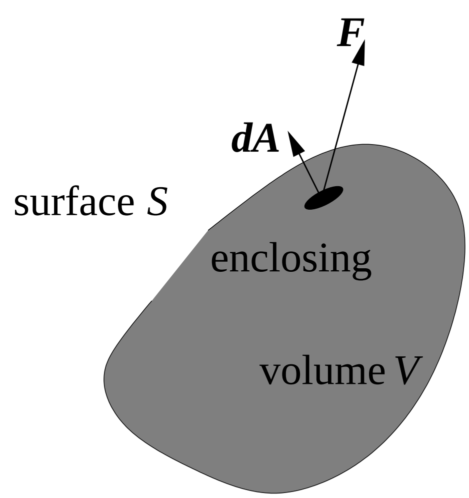
$$\begin{aligned}
\mathbf{curl}\,\vec{F} &= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right)\hat{x} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right)\hat{y} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right)\hat{z} \\
&= \vec{\nabla} \times \vec{F} .
\end{aligned}$$

Just as the divergence is nicely expressed using the gradient operator and the dot product, the curl is nicely expressed using the gradient operator and the cross product.

4.8 Wrap up and summary

That was a lot of math. Let's summarize all of these results:

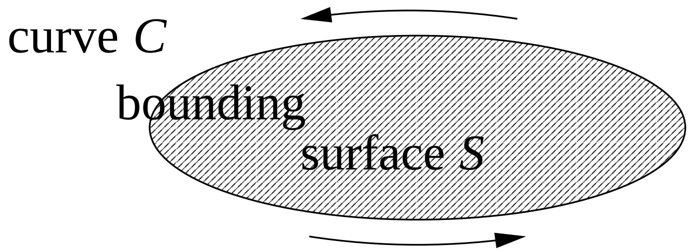
4.8.1 Gauss's theorem



$$\oint_S \vec{F} \cdot d\vec{A} = \int_V \mathbf{div}\,\vec{F} dV$$

The surface S is the boundary of the volume V.

4.8.2 Stokes's law



$$\oint_C \vec{F} \cdot d\vec{s} = \int_S \mathbf{curl}\, \vec{F} \cdot d\vec{A}\,.$$

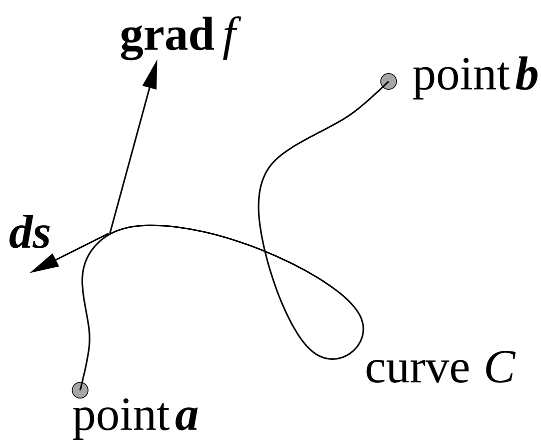
The curve C is the boundary of the surface S.

4.8.3 Gradient theorem

We only discussed this in detail for the electric potential, but the following rule holds for *any*[1] scalar function:

$$f(\vec{b}) - f(\vec{a}) = \int_{\vec{a}}^{\vec{b}} \vec{\nabla} f \cdot d\vec{s}\,.$$

This holds for any path from the point $\vec{a}$ to $\vec{b}$:



We will use these results *many* times as this course progresses.

[1]Except for some really bizarre ones which we will never encounter in this class.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 5:
FIELDS AND POTENTIALS AROUND CONDUCTORS
THE ELECTROSTATIC UNIQUENESS THEOREM

5.1 Conductors

For much of the rest of this class, we will be concerned with processes that involve *conductors*: materials in which it is easy for charges to move around. We will discuss conductors in some depth when we discuss currents; for now, we will just summarize a few of their properties.

Among the best conductors are metals — silver, gold, copper, aluminum, etc. The atoms of these metals form a crystalline structure in which electrons can easily hop around from atom to atom. Although a chunk of metal is neutral overall, we can visualize it as being made of lots of positive charges that are nailed in place, paired up with lots of negative charges (electrons) that are free to move around. In isolation, the negative charges will sit close to the positive charges, so that the metal is not only neutral overall, but also largely neutral everywhere (no *local* excess of positive or negative charge). Under the influence of some external field, the electrons are free to move around.

Some materials conduct OK, but are not as good as metals. For example, salty water has lots of charges that are free to move around under the influence of an $\vec{E}$ field. However, since these charges — usually sodium and chlorine ions — are *far* more massive than an electron, and they do not flow in a crystaline structure, salty water is not a very high quality conductor. Another example is graphite: the somewhat unusual bond structure of graphite makes it a fairly good conductor, but only in certain directions.

At the opposite end of the spectrum are *insulators*: materials in which the electrons are bound quite tightly to the constituent molecules and hence have essentially no freedom to move. Insulators are typically made from organic materials such as rubber or plastic, or from crystals formed from strong covalently bounded molecules, such as quartz or glass.

The effectiveness of a substance as a conductor is quantified by its *resistivity*, a number that expresses how well it resists the flow of current. We will revisit this quantity in about a week when we discuss currents in detail; for now, suffice to say that small resisitivity means a good conductor. In SI units, resistivity is measured in "Ohm-meters" (usually written Ω-m); in cgs units, resistivity turns out to be measured in seconds. Here are a few example values:

Material	Resistivity (Ω-m)	Resistivity (sec)
Silver	1.6×10^{-8}	1.8×10^{-17}
Copper	1.7×10^{-8}	1.9×10^{-17}
Gold	2.4×10^{-8}	2.6×10^{-17}
Iron	1.0×10^{-7}	1.1×10^{-16}
Sea water	0.2	2.2×10^{-10}
Polyethylene	2.0×10^{11}	220
Glass	$\sim 10^{12}$	$\sim 10^{3}$
Fused quartz	7.5×10^{17}	8.3×10^{8}

As you can see, the resistivity of ordinary materials varies over an enormous range, reflecting the very different electronic properties of materials around us.

5.2 Electric fields and conductors

For the rest of this lecture, we will assume that conductors are materials that have an infinite supply of charges that are free to move around. (This of course just an idealization; but, it turns out to be an extremely good one. Real conductors in fact behave very similar to this limit.) From this, we can deduce a few important facts about conductors and electrostatic fields:

- **There is no electric field inside a conductor**. Why? Suppose we bring a plus charge near a conductor. For a very short moment, there will be an electric field inside the conductor. However, this field will act on and move the electrons, which are free to move about. The electrons will move close to the plus charge, leaving net positive charge behind. The conductor's charges will continue to move until the "external" $\vec{E}$-field is cancelled out — at that point there is no longer an $\vec{E}$-field to move them, so they stay still.

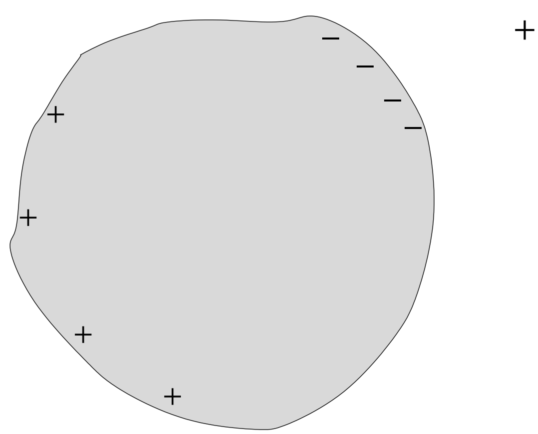


Figure 1: Conductor near an external charge. The charges in the conductor very quickly rearrange themselves to cancel out the external field.

A more accurate statement of this rule is "**After a very short time, there is no electric field inside a conductor**". How short a time is it? Recall that in cgs units, resistivity (which tells us how good/bad something conducts electricity) is measured in seconds. It turns out that the time it takes for the charges to rearrange themselves to cancel out the external $\vec{E}$-field is just about equal to this resistivity. For metals, this is a time that is something like $10^{-16} - 10^{-17}$ seconds. This is so short that we can hardly complain that the original statement isn't precise enough!

- **The electric potential within a conductor is constant**. Proof: the potential difference between any two points $\vec{a}$ and $\vec{b}$ inside the conductor is

$$\begin{aligned}\phi_b - \phi_a &= -\int_{\vec{a}}^{\vec{b}} \vec{E} \cdot d\vec{s} \\ &= 0\end{aligned}$$

  since $\vec{E} = 0$ inside the conductor. Hence, for any two points $\vec{a}$ and $\vec{b}$ inside the conductor, $\phi_b = \phi_a$.

- **Net charge can only reside on the surface of a conductor**. This is easily proved with Gauss's law: make a little Gaussian surface that is totally contained inside the conductor. Since there is no $\vec{E}$-field inside the conductor, $\oint \vec{E} \cdot d\vec{A}$ is clearly zero for your surface. Since that is equal to the charge the surface contains, there can be no charge.

  We will discuss the charge on the conductor's surface in a moment.

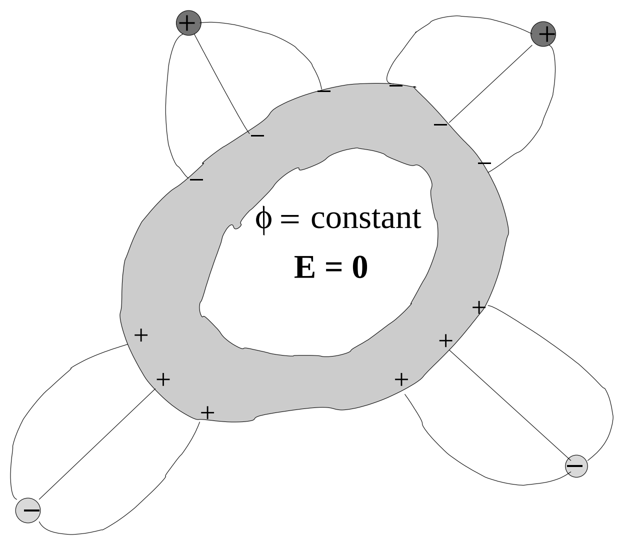
- **Any external electric field lines are perpendicular to the surface.** Another way to put this is that there is no component of electric field that is tangent to the surface. We prove this by contradiction: suppose that a component of the $\vec{E}$-field *were* tangent to the surface. If that were the case, then charges would flow along the surface. They would continue to flow until there was no longer any tangential component to the $\vec{E}$-field. Hence, this situation cannot exist: even if it exists momentarily, it will rapidly (within 10^{-17} seconds or so) correct itself.

- **The conductor's surface is an equipotential.** This follows from the fact that the $\vec{E}$-field is perpendicular to the surface. We do a line integral of $\vec{E}$ on the surface; the path is perpendicular to the field; so the difference in potential between any two points on the surface is zero.

A few important corollaries follow from these rules.

- **Corollary 1**: Consider a conductor with a hollowed out region. If there is no charge in this hollow, then the potential ϕ there is constant. It follows that the electric field $-\vec{\nabla}\phi$ inside the hollow is zero.

  We will prove this carefully shortly. For now, we can motivate this proof by noting that the surface of the conductor must be an equipotential. Since there is no charge anywhere on the inside, the interior potential must obey Laplace's equation, $\nabla^2\phi = 0$. Solutions to Laplace's equation can have no local maxima or minima. The only solution that has some proscribed constant value on an exterior boundary *and* has no local maxima or minima is one that is constant.



By putting things inside a conducting box, it is shielded from any external electric field. Such an arrangement is known as a Faraday cage. Although it is beyond the scope of this course to prove this, Faraday cages work pretty well even when the external fields vary in time, as long as they don't vary too quickly.

- **Corollary 2**: suppose we put a charge q inside this hollow. There must be an *induced* layer of surface charge on the wall of the hollow; the total amount of induced charge in this layer is $q_{\rm ind} = -q$.

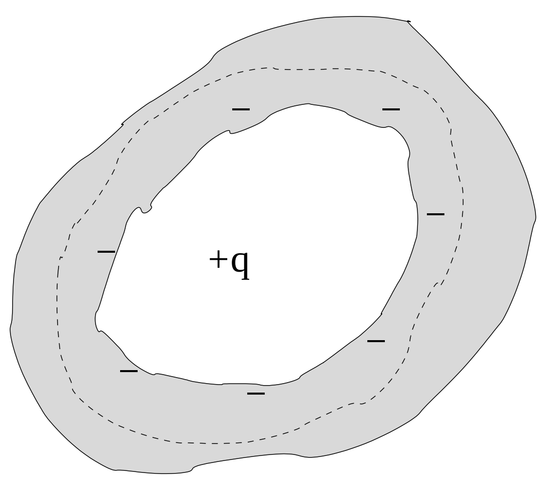


Figure 2: Gaussian surface: dashed line enclosing hollow and induced negative charge.

To prove this, use Gauss's law. We draw a Gaussian surface that lies inside the conductor. The $\vec{E}$-field there is zero, so the total electric flux must be zero. This flux equals 4π times the total charge contained by the surface, which is q plus the induced charge:

$$\begin{aligned} \oint \vec{E}\cdot d\vec{A} &= 0 \\ &= 4\pi(q + q_{\rm ind}) \\ \rightarrow q_{\rm ind} &= -q\ . \end{aligned}$$

You should be able to prove with another well chosen Gaussian surface that the conductor has another layer of charge on its *outer* surface with total magnitude $+q$.

- **Corollary 3**: The induced charge *density* on any surface is given by $|\vec{E}|/4\pi$, where $\vec{E}$ is the electric field right next to the surface. This follows from the rule that whenever there is a surface charge layer, the electric field changes with $\Delta E = 4\pi\sigma$. Since the field inside the conductor is zero, $\Delta E = |\vec{E}|$.

5.3 The uniqueness theorem

Suppose we know that some region contains a charge density $\rho(\vec{r})$. Suppose we also know the value of the electronstatic potential $\phi(\vec{r})$ value on the boundary of this region. The *uniqueness theorem* then guarantees that there is only one function $\phi(\vec{r})$ which describes the potential in that region. This means that no matter how we figure out ϕ's value — guessing, computer aided numerical computation, demonic invocation — the function ϕ we find is *guaranteed* to be the one we want.

We prove this theorem by assuming it is not true: we assume that two functions, $\phi_1(\vec{r})$ and $\phi_2(\vec{r})$, *both* satisfy Poisson's equation:

$$\begin{aligned} \nabla^2\phi_1(\vec{r}) &= -4\pi\rho(\vec{r}) \\ \nabla^2\phi_2(\vec{r}) &= -4\pi\rho(\vec{r}) \,. \end{aligned}$$

Both of these functions must satisfy the boundary condition: we must have $\phi_1(\vec{r}) = \phi_2(\vec{r}) = \phi_B(\vec{r})$ on the boundary of our region.

By the principle of superposition, any combination of the potentials $\phi_1(\vec{r})$ and $\phi_2(\vec{r})$ must be a perfectly valid potential. (It's not the potential that describes our region, but it still is valid as far as the general laws of physics are concerned.) Let's look in particular at $\phi_3(\vec{r}) = \phi_2(\vec{r}) - \phi_1(\vec{r})$. First, what is the boundary condition for ϕ_3? Since $\phi_1 = \phi_2$ on the boundary, we must have $\phi_3 = 0$ there. Next, apply the Laplacian to ϕ_3:

$$\begin{aligned} \nabla^2\phi_3 &= \nabla^2\left(\phi_2 - \phi_1\right) \\ &= \nabla^2\phi_2 - \nabla^2\phi_1 \\ &= 4\pi\rho - 4\pi\rho = 0 \,. \end{aligned}$$

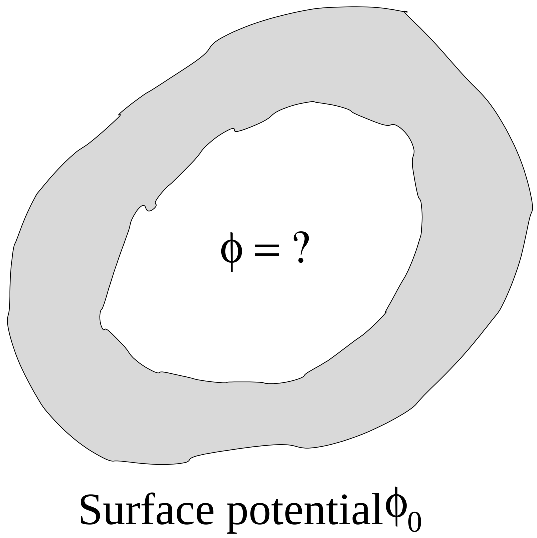
The potential ϕ_3 thus satisfies Laplace's equation. This means that it can have no local maxima or minima inside its boundary. But, on the boundary, its value is zero! The only function which is zero on a boundary **and** has no local maxima or minima is one which is zero *everywhere* in the region: $\phi_3(\vec{r}) = 0$. This means that $\phi_1(\vec{r}) = \phi_2(\vec{r})$ — our initial assumption, that at least two potentials satisfied Poisson's equation in the region with the given boundary condition, was wrong. Hence the solution for $\phi(\vec{r})$ is unique. Putting it concisely, we have proved the uniqueness theorem: **An electrostatic potential $\phi(\vec{r})$ is completely determined within a region once its value is known on the region's boundary.**

This is good news: it gives us license to be lazy.

5.4 Some conductor examples

5.4.1 Example 1: Randomly shaped conductor

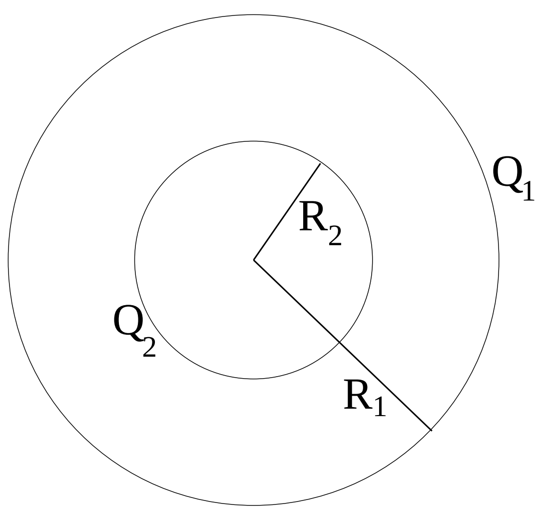
A chunk of conductor with a hollowed out core has charge placed on it until its potential (relative to infinity) is ϕ_0. What is the potential inside the core?



Answer: the potential inside the core — indeed, everywhere inside the conductor's surface — must be ϕ_0. Why? $\phi(\vec{r}) = \phi_0$ is a solution that satisfies the boundary condition, and it obviously satisfies Laplace's equation (since ϕ_0 is a constant). By the uniqueness theorem, that's all we need.

5.4.2 Example 2: Nested concentric spherical shells

A pair of thin, conducting, nested spherical shells with radii R_2 and R_1 carry charges Q_2 and Q_1 respectively. What is the potential of the inner sphere? What is the potential of the outer sphere? What is the potential at $r = 0$? (Set the zero of the potential at infinity.)



Let's do the potential of the outer sphere first: it must be

$$\phi_1 = \frac{Q_1 + Q_2}{R_1} \, .$$

This should hopefully be almost obvious — if not, apply Gauss's law outside the outer sphere and note that it looks just like a point charge of $Q_1 + Q_2$ at the origin.

To get the potential of the inner sphere, we use

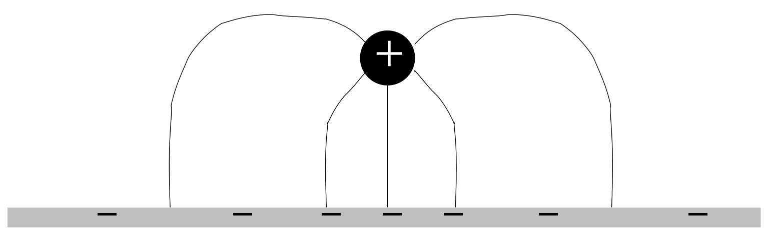
$$\begin{aligned}
\phi_2 - \phi_1 &= -\int_{R_1}^{R_2} \vec{E} \cdot d\vec{s} \\
&= -\int_{R_1}^{R_2} \frac{Q_2}{r^2} dr \\
&= \frac{Q_2}{R_2} - \frac{Q_2}{R_1} \\
\to \phi_2 &= \frac{Q_2}{R_2} + \frac{Q_1}{R_1} \, .
\end{aligned}$$

By the uniqueness theorem, the potential at $r = 0$ — indeed, for *any* $r \leq R_2$ — must be ϕ_2.

5.4.3 Example 3: Point charge and flat plane conductor

.

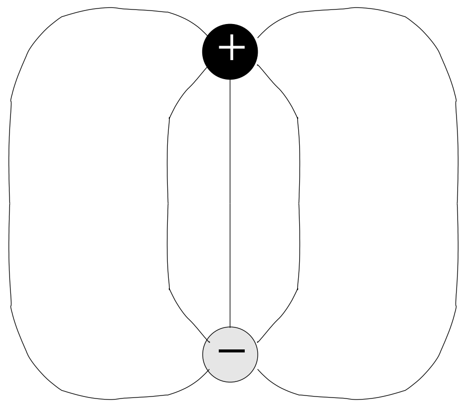
Suppose we hold a point charge Q a distance d above an infinite, flat plane conductor. The potential on the conductor is fixed at zero[1]. What is the potential everywhere above the conductor?



A priori, this looks nasty: the point charge induces minus charge on the conductor, which makes things really complicated. What do we do?

[1]This is easily done by hooking it up to *ground*. "Ground" effectively means an infinite supply of plus and minus charges. If we define ground as our reference potential, $\phi = 0$, then anything hooked up to it will also be at $\phi = 0$.

Uniqueness theorem to the rescue. It doesn't matter how we find the potential, as long as we do so in a way that satisfies the boundary condition: the potential must be $\phi = 0$ on the conductor, and the electric field must be purely vertical in the conductor's plane. A *different* configuration which produces the *exact same potential* above the plane is the following one:



We've eliminated the plane entirely, and introduced a minus charge a distance d *below* the conductor's surface. This fictitious charge is called an *image* charge. With this configuration, the plane right between the two charges (where the conductor's surface is located) will be at $\phi = 0$ (superposition: each point is equidistant from both charges, which are of opposite sign). Also the field is purely vertical at that plane (horizontal components cancel by symmetry). The potential satisfies the boundary condition; it satisfies Laplace's equation; therefore, by the uniqueness theorem it is **THE** solution.

The solution is of course complete nonsense in the region below the conductor's surface. But we don't care! That region was not part of the study anyway.

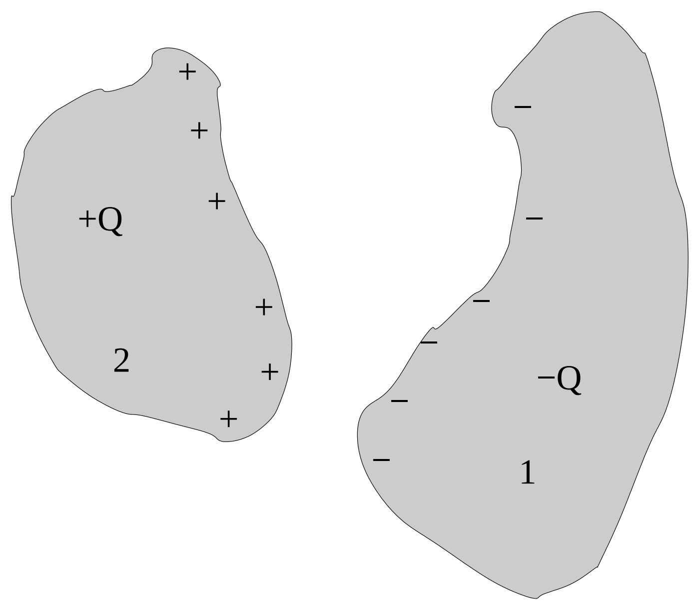
Scott Hughes *17 February 2005*

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 6:
Capacitance

6.1 Capacitance

Suppose that I have two chunks of metal. A charge $+Q$ is on one of these chunks, $-Q$ is on the other (so that the system is neutral overall). Each chunk will be at some constant potential. What is the potential difference between the two chunks of metal?



Whatever the potential difference $V \equiv \phi_2 - \phi_1 = -\int_1^2 \vec{E} \cdot d\vec{s}$ turns out to be, it must of course turn out to be independent of the integration path. In fact, as we can show with a little thought, the potential difference V must be proportional to the geometry. This means that the potential difference (or "voltage") between the two chunks must take the form

$$V = (\text{Horribly messy constant depending on geometry}) \times Q\ .$$

The horrible mess that appears in this proportionality law depends only on geometry. In other words, it will depend on the size and shape of the two metal chunks, their relative orientation, and their separation. It does *not* depend on their charge.

This constant is defined as $1/C$, where C is the *Capacitance* of this system. A system like this is then called a *capacitor*. The 1/ may seem a bit wierd; the point is that one usually writes the capacitance formula in a different way: we put

$$Q = CV\ .$$

A key thing to bear in mind when using this formula is that Q refers to the charge *separation* of the system. In other words, when I say that a capacitor is charged up to some level Q,

Figure 1: $Q = CV$.

I mean that I have put $+Q$ on one part of the capacitor, $-Q$ on the other. *The capacitor as a whole remains neutral!!* I emphasize this now because I often find students are somewhat confused by this point, particularly when we start thinking about circuits that have capacitors in them.

6.1.1 Why are the charge and voltage proportional?

The real concern here is, how do we know that when we say double the charge on the capacitor, the charge doesn't move around and cause the dependence to be more complicated? The simple answer is the principle of superposition: whenever we double the amount of charge, the fields that these charges create simply double. Hence, the potential difference (which is just the line integral of the field) must also double.

There's an implied assumption here, though: we are assuming that there is enough charge smeared out on the "plates" of the capacitor to uniformly cover them. Imagine we just had one electron on the minus plate, and one electron "hole" on the plus plate. If we just doubled the charge, the electron on the - plate indeed might move somewhere else. This situation could be messy!

Fortunately, in every situation that is of practical interest, the assumption that the plates are evenly coated with a charge excess is a good one. This is because the elementary charge is so small — even if there's just a tiny tiny amount of charge — say 10^{-12} Coulombs — that means we've got something like 6 million excess electrons.

6.1.2 Units of capacitance

In SI units, we measure charge in Coulombs and potential in Volts, so the unit of capacitance is the Coulomb/Volt. This combination is given a special name: the Farad, or F.

In cgs units, we measure charge in esu and potential in esu/cm. The unit of capacitance is thus just the centimeter! This is actually kind of cool: objects that are big (have lots of centimeters) tend to have a big capacitance.

It's important to know how to convert between these two units: when we are computing capacitance from the system's characteristics, cgs units are nice, since it's just a matter of getting the geometry right. But SI units are used almost exclusively in circuits.

$$\begin{aligned} 1\,\mathrm{cm} &= 1.11 \times 10^{-12}\,\mathrm{F} \\ &\simeq 1\,\mathrm{pF}\,. \end{aligned}$$

"pF" means "picofarad" — one trillionth of a Farad. Capacitors typically come in microfarads or picofarads; this is because (as we shall see in a moment) a Farad turns out to be an enormous amount of capacitance.

6.2 Some examples

6.2.1 Isolated sphere

We take a sphere of radius R and put a charge Q on it. Its potential (relative to infinity) is $V = Q/R$. The sphere's capacitance is therefore

$$C_{\mathrm{sphere}} = R\,.$$

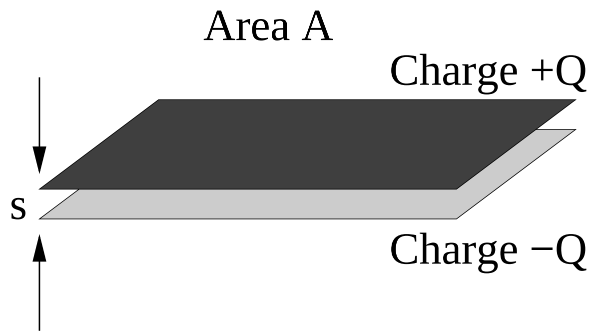
When we began discussing the notion of capacitors, we talked about the potential difference between two conductors. In this case, the second conductor is a "virtual" plate at infinity.

There's one spherical capacitor that we use all the time: the earth. It has a radius (and hence a capacitance) of 6.4×10^8 cm. This is so large that we can effectively take charge from or dump charge onto the earth without changing its electrical potential at all. Converting, the earth has a capacitance $C_{\mathrm{earth}} = 0.0007$ Farad — enormous, but still significantly smaller than a Farad!

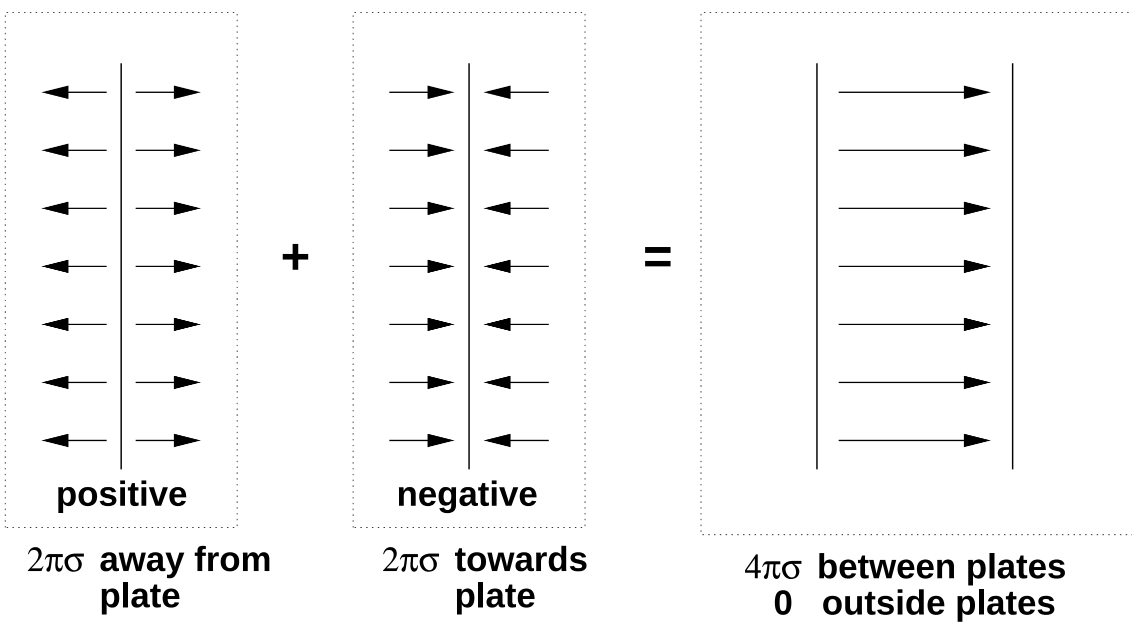
The fact that we can dump charge onto or take charge from the earth without changing its potential appreciably means that it is often convenient to define the earth as the reference point for computing potential. Many circuits take advantage of this by hooking directly to "ground" — a lead that runs into the earth. That "grounded lead" is then defined, by convention, as potential zero. (Students educated in the remnants of the British Empire may have learned to call a circuit that is grounded "earthed".)

6.2.2 Parallel plates

Consider a pair of plates, each with area A and separated by a distance s.



If the plates are close together ($s \ll \sqrt{A}$), we can approximate the field between these plates as those coming from a pair of infinite planes:



Summing an infinite plane with charge density $\sigma = +Q/A$ and $\sigma = -Q/A$ we find

$$E = 4\pi\sigma = 4\pi Q/A$$

between the plates, and zero outside of the plates. The field points from the positive plate to the negative plate, so the potential difference from the negative to the positive plate is

$$V = -\int_{\text{neg}}^{\text{pos}} \vec{E} \cdot d\vec{s} = 4\pi\sigma \int_0^s ds = \frac{4\pi Q s}{A} \ .$$

Rewriting this in the form $Q = CV$, we read off the capacitance:

$$C_{\text{plates}} = \frac{A}{4\pi s} \ .$$

This is an area divided by a length, so it has the correct units.

It's pretty important to know about capacitance in SI units, since most circuit elements are discussed using things like Volts and Farads rather than statvolts and centimeters. Converting is easy: we just need to remember that the electric field in general has a factor of k (the constant from Coulomb's law) attached to it. Hence, the voltage V likewise picks up this factor; rearranging to solve for the capacitance, we find

$$C_{\text{plates}} = \frac{1}{k}\frac{A}{4\pi s}\ .$$

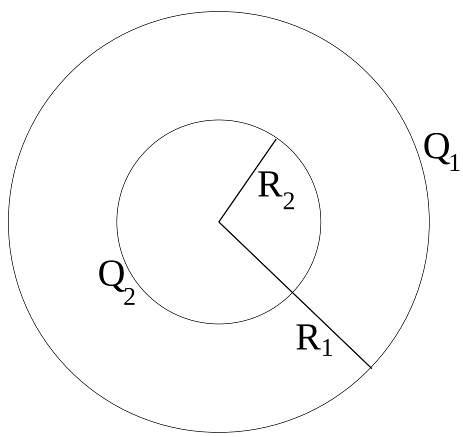
In cgs units, $k = 1$, so this rather trivially reduces to what we had before. In SI units, $k = 1/4\pi\epsilon_0$, and we find

$$C_{\text{plates}} = \frac{\epsilon_0\, A}{s}\ .$$

Whenever you work out a formula for capacitance in cgs units, you can easily convert to SI by multiplying by $1/k = 4\pi\epsilon_0$.

6.2.3 Nested shells

Return to the example discussed last lecture: a pair of nested spherical shells,



The potential difference we found in this case was

$$\Delta\phi = V = \frac{Q_2}{R_2} - \frac{Q_2}{R_1}\ .$$

Set $Q_2 = Q = -Q_1$:

$$\Delta\phi = V = Q\left(\frac{1}{R_2} - \frac{1}{R_1}\right)\ .$$

Rearranging into the form $Q = CV$, we find

$$C = \frac{R_1 R_2}{R_1 - R_2}\ .$$

Note that if R_1 is just barely larger than R_2, this gives the same result as the parallel plate formula: putting $R_1 = R_2 + s$, we find

$$\begin{aligned} C &= \frac{R_2^2 + R_2\, s}{s} \\ &\simeq \frac{R_2^2}{s} \qquad s \ll R_2 \\ &= \frac{4\pi R_2^2}{4\pi s} \\ &= \frac{A}{4\pi s} \,. \end{aligned}$$

On the second to last line, we've used the fact that $4\pi R_2^2$ is the surface area of the sphere.

6.3 Energy stored in a capacitor

Suppose we are charging up a capacitor. At some intermediate point in the process, there is a charge $+q$ on one plate, $-q$ on the other. We move a charge dq from the negative charge to the positive charge. It takes work to do this: we are forcing a positive charge in the direction opposite to what it "wants" to do. The amount of work we do moving this charge is

$$\begin{aligned} dW &= V(q)\, dq \\ &= \frac{q}{C} dq \,. \end{aligned}$$

Integrating this to a total charge separation of Q, we find that the work that is done charging up the capacitor is

$$\begin{aligned} W &= \int_0^Q \frac{q}{C} dq \\ &= \frac{1}{2}\frac{Q^2}{C} \,. \end{aligned}$$

This is the energy stored that is stored in the capacitor:

$$U = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2} C V^2 \,.$$

We've used $Q = CV$ in the second equality.

It's useful to carefully re-examine this formula for the case of the parallel plate capacitor: we put $C = A/4\pi s$, and note that the charge $Q = EA/4\pi$, where E is the electric field between the plates. Then,

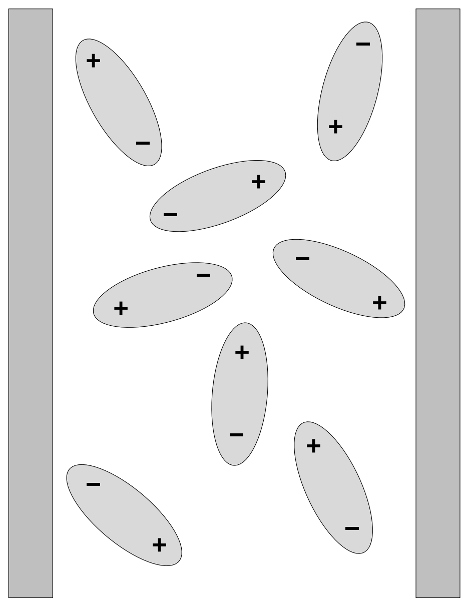
$$\begin{aligned} U &= \frac{1}{2}\frac{Q^2}{C} \\ &= \frac{1}{2}\frac{E^2 A^2}{16\pi^2}\frac{4\pi s}{A} \\ &= \frac{E^2}{8\pi} A s \,. \end{aligned}$$

The last line is the energy density of the electric field between the plates times the volume of that region.

6.4 Dielectrics

The capacitors we've discussed here are somewhat artificial: the space between their plates is taken to be empty. In the real world, the normal situation is that this space is filled with something — air, plastic, paper, compressed leprechauns, whatever. The "stuff" that fills this space interacts with the electric field, and can have an important impact on the the properties of the capacitor.

The basic idea can be understood fairly simply. The key point is that many materials are *polarizable*: they are made out of molecules that, though neutral overall, have a net excess of + charge at one end and an excess of − charge at the other. This means that each molecule looks like a little electric "dipole" — a distribution whose total charge is zero, but that shows non-trivial charge separation. Normally, the polar character of these molecules is unimportant — the random, mostly thermal motion of the molecules means that at any moment there is no net polarization to a big mass of them. If we take a block of such material and slide it between two parallel plates, it looks (roughly) like this:

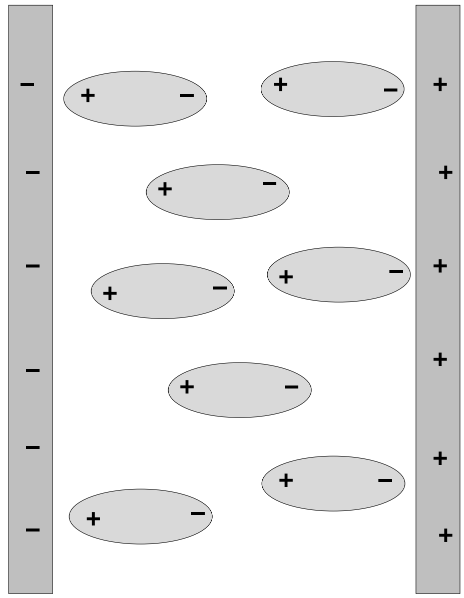


Suppose we now charge up these plates, so that there is an electric field between them. We leave these plates with *fixed charge*. All of these little dipoles will now want to line up — the + ends will align with the negatively charged plate, and vice versa:

By lining up in this way, they will tend to *reduce* the electric field between the plates — the field due to the charges on the plates themselves is superposed upon a field pointing in the opposite direction due to the alignment of the dipoles. For an enormous number of molecules, the reduction in the applied electric field can be summarized using a single number:

$$\text{Fixed charge situation:} \qquad E^{\text{with dielectric}} = E^{\text{without dielectric}}/K\ .$$

The number K is called the *dielectric constant*; it's something we can measure and catalog for various substances. Table 10.1 of Purcell lists a whole bunch; the values range from $K \simeq 1$ for air to $K \simeq 80$ for water. (Water has a huge dielectric constant because of the charge distribution associated with its bent molecules.) For us, the value of K is always greater than or equal to one.



Since the electric field is reduced by K, the voltage between the plates is likewise reduced. For the parallel plate capacitor and in cgs units, we have

$$V^{\text{with dielectric}} = \frac{4\pi Q s}{KA} \; .$$

Invoking the definition of the capacitance, we find

$$C^{\text{with dielectric}} = \frac{KA}{4\pi s} = KC^{\text{without dielectric}} \; .$$

It's worth thinking for a moment about what the dielectric does in some specific circumstances. Suppose I have a capacitor C whose plates have been charged up to some *fixed charge* Q. This is easily done by hooking a capacitor up to a voltage source V (e.g., a battery) and then removing the source — the capacitor will then hold the charge $Q = CV$. If I insert a dielectric between the plates, the capacitance increases; since the charge is fixed, the voltage must decrease:

$$\text{Fixed charge:} \qquad V^{\text{new}} = \frac{Q}{C^{\text{new}}} = \frac{Q}{KC^{\text{orig}}} = \frac{V^{\text{orig}}}{K} \; .$$

Suppose instead I take the capacitor and hold it at *fixed voltage*. This is also easy to do — we hook up our capacitor to a voltage source and just leave it there. The capacitance again increases, which means the charge must increase:

$$\text{Fixed voltage:} \qquad Q^{\text{new}} = C^{\text{new}} V = KC^{\text{orig}} V = KQ^{\text{orig}} \; .$$

The "extra" charge is actually pulled from the voltage source. Note that *in this circumstance* the electric field between the plates does not decrease when we insert the dielectric! However, it takes *more charge* to hold this field than it does without the dielectric — the dielectric "wants" to reduce the electric field; the voltage source needs to supply more charge to "fight" this tendency.

To wrap this section up, consider the capacitance in SI units. With a dielectric, it is given by

$$
\begin{aligned}
C^{\text{with dielectric}} &= \frac{K\epsilon_0 A}{s} \\
&\equiv \frac{\epsilon A}{s} \,.
\end{aligned}
$$

The combination $K\epsilon_0 \equiv \epsilon$ is called the *permittivity* of the dielectric. When $K = 1$, the permittivity is just ϵ_0, the permittivity of "free space" (meaning "empty space").

In practice, people don't usually describe dielectrics using permittivity — the dielectric constant K is much more useful. However, you do encounter the term sometimes; and, it helps to explain why ϵ_0 is called the permittivity of free space.

Scott Hughes *24 February 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

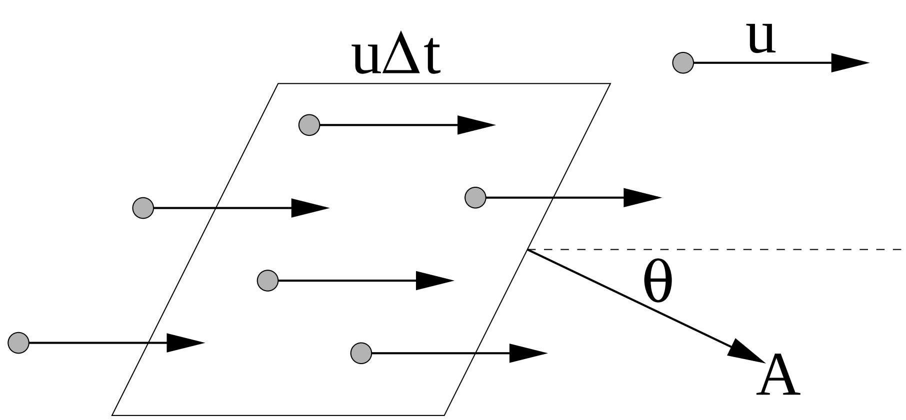
LECTURE 7:
CURRENT, CONTINUITY EQUATION, RESISTANCE, OHM'S LAW.

7.1 Electric current: basic notions

The term "electric current" is used to describe the charge per unit time that flows through a region. In cgs units, current is measured in esu/sec, naturally enough. In SI units, current is measured in Coulombs/sec, which is given the name *Ampere* (or amp). It is very important to know how to convert between these units! Amps are used to describe currents *FAR* more often than esu/sec: 1 ampere $= 2.998 \times 10^9$ esu/sec.

Suppose we have a swarm of charges, all with the same charge q. The *number density* of these charges is some value n (i.e., there are n charges per unit volume). Suppose further that all of these charges are moving with velocity $\vec{u}$. How much current is flowing through an area $\vec{A}$?

To figure this out, we need to calculate how many of the charges pass through the area $\vec{A}$ in time Δt. This number is given by the number of charges that fit into the oblique prism sketched below:



The number of charges in this prism is its volume, $|\vec{A}||\vec{u}\Delta t|\cos\theta$, times the number density. The number of charges that pass through the area in time Δt is

$$\Delta N = n(|\vec{u}|\Delta t)(|\vec{A}|)\cos\theta = n\vec{u}\cdot\vec{A}\Delta t\ .$$

The total *charge* that passes through is q times this number:

$$\Delta q = qn\vec{u}\cdot\vec{A}\Delta t\ ,$$

so that the current $I = \Delta q/\Delta t$ must be

$$I = qn\vec{u}\cdot\vec{A}\ .$$

This form of the current motivates the definition of the *current density* $\vec{J}$:

$$\vec{J} = qn\vec{u} = \rho\vec{u} \; .$$

For this (admittedly artificial) example of all charges streaming in the same direction with the same velocity $\vec{u}$, the current density is just the charge density $\rho \equiv nq$ times that velocity. The current is then given by $I = \vec{J} \cdot \vec{A}$.

Caution: bear in mind that the current density is a current *per unit area*: charge/(time $\times$ area). This is hopefully obvious by dimensional analysis: charge density is charge/(length cubed); velocity is length/time; hence current density is charge/(time $\times$ length squared).

7.2 Electric current: more details

Some of the details assumed above are obviously fairly artificial. We make them more realistic one by one. First, rather than having a single kind of charge that is free to move, a material might have a bunch of different charges that can carry current. For example, ocean water conducts electricity largely because of the freely moving sodium and chlorine ions. Other materials dissolved in the ocean contribute a bit as well. Each variety of charge may have its own charge, number density, and velocity. Let the subscript k label different charge "species" — e.g., $k = 1$ could refer to electrons, $k = 2$ to chlorine ions, etc. The total current density comes from combining them all together:

$$\vec{J} = \sum_k q_k n_k \vec{u}_k = \sum_k \rho_k \vec{u}_k \; .$$

Note, via this formula, the equivalence between positive and negative charges we get by flipping the velocity direction: a positive charge moving with $\vec{u}_k$ contributes the same amount to the overall current density as a negative charge moving with $-\vec{u}_k$.

Next, the assumption that *all* charges of species k are moving with the exact same velocity $\vec{u}_k$ is pretty obviously ludicrous. A current of 1 amp corresponds to (for example) 6.24×10^{18} electrons per second streaming past. There's no way every single one of those electrons has the **exact** same velocity! So, in our formula for $\vec{J}$, we replace $\vec{u}_k$ with its value averaged over all the charges:

$$\langle \vec{u}_k \rangle = \frac{1}{N_k} \sum_{i=1}^{N_k} (\vec{u}_k)_i \; ,$$

where N_k is the total number of charge carriers of species k, and $(\vec{u}_k)_i$ is the velocity of the ith charge in species k (admittedly, not the sanest looking equation in the universe). The current density is thus given by

$$\vec{J} = \sum_k q_k n_k \langle \vec{u}_k \rangle = \sum_k \rho_k \langle \vec{u}_k \rangle \; .$$

(Note: Purcell gives a somewhat different formula for $\vec{J}$ built from the averaged velocity. I find his formula to be rather confusing; it took me half an hour to connect what he does with my own understanding of current density and averages. De mortuis nil nisi bonum.)

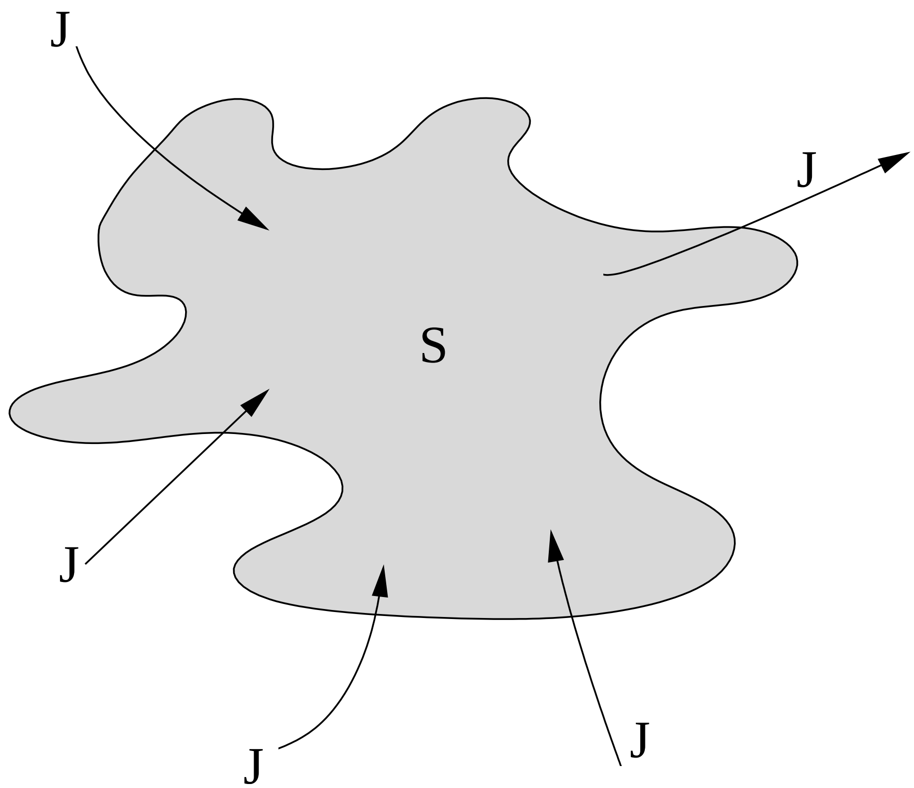
Finally, we do not expect in general to encounter nice, simple, uniform surfaces through which our current flows. Not surprisingly, we generalize to an integral: the current — charge/time — passing through some general surface S is given by

$$I = \int_S \vec{J} \cdot d\vec{A} \; .$$

We can thus think of total current as the flux of current density through a surface.

7.3 Charge conservation and continuity

Suppose the current flows through a *closed* surface:



In this case, there *must* be charge piling up inside (or departing from) the volume V enclosed by the surface S:

$$\oint_S \vec{J} \cdot d\vec{A} = -\frac{\partial}{\partial t} Q(\text{contained by } S)\ .$$

The minus sign in this equation comes from our convention on assigning direction to the area element $d\vec{A}$ — since $d\vec{A}$ points out, $\vec{J} \cdot d\vec{A}$ is negative if the current density points inward, and is positive if the current density points outward. Since inward corresponds to a gain of charge (and vice versa) we clearly need a minus sign for the physics to make sense.

Using our skill at vector calculus, we can manipulate this equation as follows:

$$\oint_S \vec{J} \cdot d\vec{A} = \int_V \left(\vec{\nabla} \cdot \vec{J}\right) dV\ ,$$

where V is the volume enclosed by S. This is just Gauss's theorem. We also have

$$Q(\text{enclosed by } S) = \int_V \rho\, dV\ .$$

Putting this together we have

$$\begin{aligned}
\int_V \left(\vec{\nabla} \cdot \vec{J}\right) dV &= -\frac{\partial}{\partial t} \int_V \rho\, dV \\
&= -\int_V \frac{\partial \rho}{\partial t} dV \\
\to \int_V \left(\vec{\nabla} \cdot \vec{J} + \frac{\partial \rho}{\partial t}\right) dV &= 0\ .
\end{aligned}$$

This equation is guaranteed to hold only if

$$\vec{\nabla} \cdot \vec{J} + \frac{\partial \rho}{\partial t} = 0 \; .$$

This result is known as the *continuity equation*: it guarantees conservation of electric charge in the presence of currents. It also tells us that when currents are *steady* — no time variation, so that $\partial\rho/\partial t = 0$ — we must have $\vec{\nabla} \cdot \vec{J} = 0$.

7.4 Ohm's law

Electric fields cause charges to move. It stands to reason that an electric field applied to some material will cause currents to flow in that material.

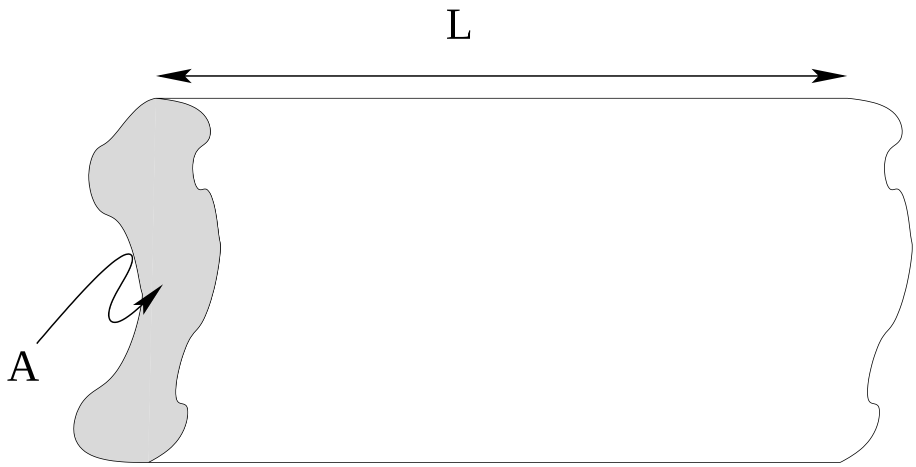
For an amazingly wide range of materials, an empirical rule called Ohm's law gives the following relation between current density and applied electric field:

$$\vec{J} = \sigma \vec{E} \; .$$

In other words, the current density is directly proportional to the electric field. The constant of proportionality σ is called the material's *conductivity*.

CAUTION: the symbol σ stands for BOTH conductivity AND surface charge density!!!! This can be annoying, but is a convention that we are stuck with, sadly. You should always know from the context of a given equation the physical meaning of a particular σ. Even more sadly, there are times when both meanings of σ need to be used simultaneously. In such a case, I suggest using subscripts to distinguish the two symbols — e.g., σ_c gives the conductivity, σ_q is surface charge density.

Ohm's law comes in two flavors. The version already given, $\vec{J} = \sigma\vec{E}$, can be thought of as the "local" or "microscopic" version. It tells me about the relation between current density and field at *any given point* in some material. To understand the second version, consider current flowing in some chunk of material. The material is perfectly homogeneous, meaning that the conductivity is σ throughout the chunk. It is of length L, and has a cross-section area A:



We put this material in a *uniform* electric field $\vec{E}$. Let's orient this field parallel to the material's long axis. Then, the potential difference or "voltage" between its two ends is $V = EL$.

At every point in the chunk's interior, Ohm's law holds: $J = \sigma E$ (dropping the vector signs since everything points along the long axis). Since E is constant everywhere, and since the chunk is homogeneous, this value of J holds everywhere. The *total* current flowing past any point must be $I = JA$. We now look for a relationship between the current I and the voltage V:

$$\begin{aligned} J &= \sigma E \\ \frac{I}{A} &= \sigma\frac{V}{L} \\ \rightarrow V &= I\left(\frac{L}{\sigma A}\right) \\ &\equiv IR\,. \end{aligned}$$

The formula $V = IR$ is the "global" or "macroscopic" form of Ohm's law: it is entirely equivalent to $J = \sigma E$, but involves quantities which are integrated over the macroscopic extent of our conducting material. The quantity in parentheses is defined as the *resistance* R of the chunk of material:

$$R \equiv \frac{L}{\sigma A} \equiv \frac{\rho L}{A}\,.$$

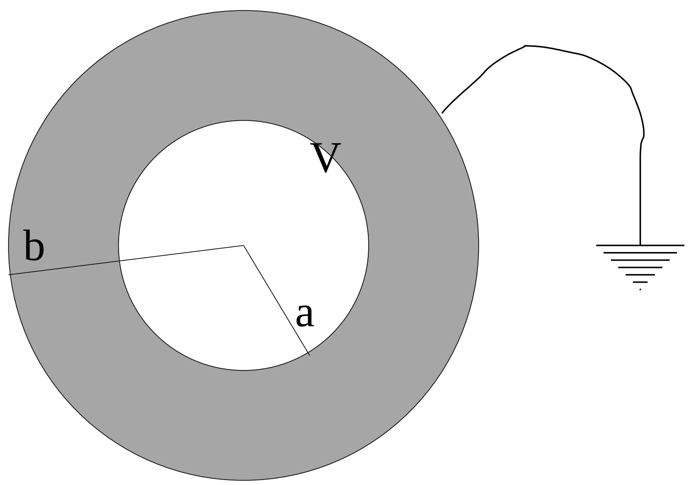
In this context, the symbol $\rho \equiv 1/\sigma$ is the *resistivity* of the material. Sadly, this is yet another example of overusing symbols — be careful not to confuse ρ as resistivity with ρ as charge density. The key thing to note about the resistance is that it goes proportional to the length of a material, and inversely proportional to the cross-sectional area.

Units: In SI units, resistance is measured in Ohms, which is just a special name for Volts/Amperes. In cgs units, the voltage is measured in esu/cm, and the current in esu/sec, so the resistance is measured in sec/cm, a rather odd unit. From this, we can infer the units of resistivity and conductivity: resistivity is measured in Ohm-meters in SI units[1], and in seconds in cgs.

7.4.1 Example: Resistance of a spherical shell

We derived the formula for the resistance of a piece of material for a very simple geometry. Similar calculations apply even if things are arranged in a goofier way.

Suppose we have a pair of nested spheres; the space between them is filled with material whose resistivity is ρ. The inner sphere is held at constant potential V; the outer sphere is *grounded*, so that its potential is zero. The radius of the inner sphere is a; that of the outer sphere is b. The current flows steadily — there is no charge piling up anywhere. What is the resistance of this shell?



We want to use Ohm's law in microscopic form, $J = \sigma E$, express the field in terms of V, the current density in terms of a current I, and then read off the resistance R. To do that, we need to know the potential at any point between the plates. By spherical symmetry, the potential must clearly have the form $\phi = C/r + D$, for some constants C and D. We have the boundary conditions that at $r = a$, $\phi = V$; at $r = b$, $\phi = 0$:

$$\begin{aligned} V &= \frac{C}{a} + D \\ 0 &= \frac{C}{b} + D \,. \end{aligned}$$

With a bit of effort, we solve these equations to find

$$\begin{aligned} C &= V\frac{ab}{b-a} \\ D &= -V\frac{a}{b-a} \,. \end{aligned}$$

[1] In practice, this unit is rarely used. The unit Ohm-cm is used much more often.

Now, we use $E = -\partial\phi/\partial r$ to find

$$E = \frac{V}{r^2}\frac{ab}{b-a} .$$

So, we now know the electric field. Note this points radially.

How much current passes through a shell of radius r somewhere between a and b? We must have

$$I = 4\pi r^2 J .$$

Note that this current must be constant! There is no charge piling up, so the same total charge per unit time passs through any given shell of radius r.

Now, we impose the microscopic form of Ohm's law:

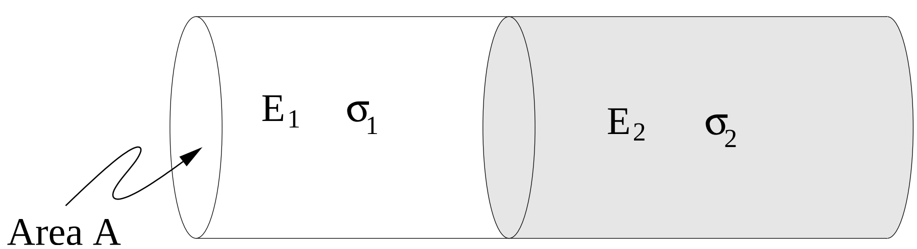
$$\begin{aligned} I &= (4\pi r^2)(\sigma E) \\ &= (4\pi r^2)\sigma\frac{V}{r^2}\frac{ab}{b-a} \\ &= \frac{4\pi\sigma ab}{b-a}V . \end{aligned}$$

Rearranging to put this in the form $V = IR$, putting $\rho = 1/\sigma$, we read off the resistance of the shell:

$$R = \frac{\rho(b-a)}{4\pi ab} .$$

7.4.2 Example: Variation of σ

What happens if σ is not a constant? The answer is set by requiring that the current flow be steady. Consider two conductors of conductivity σ_1 and σ_2 fused together:



The current must be the same in each half:

$$I = JA = \sigma_1 E_1 A = \sigma_2 E_2 A .$$

This can only be true if the electric field is different in each material. The electric field thus suddenly jumps as we cross from one side of the junction to the other, which means that there must be a layer of surface charge σ_q at the boundary[2]:

$$\sigma_q = \frac{1}{4\pi}(E_2 - E_1) = \frac{I}{4\pi A}\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right) .$$

Whenever the conductivity varies, there is a possibility that, in the steady state, some charge accumulates somewhere. This is necessary to maintain steady current flow.

[2]Here's where the overuse of the symbol σ really gets annoying. Sorry.

7.5 A closer look at Ohm's law

Ohm's law, on the face of it, makes no sense. It tells us that under an applied electric field, we get a steady current density. Since $\vec{J} = qn\vec{u}$ (considering only one current carrier), this means that the electric field is causing currents to move at constant velocity.

This should bug you! Electric fields cause *forces*; forces cause *accelerations*. Why don't the charges just accelerate like mad?

The answer is that the charges are in a medium — the conductor — that prevents them from moving freely. The motion is essentially like that in a viscous medium. The situation is much closer to a sky diver in free fall — after a short period of acceleration, the charges reach "terminal velocity", and essentially ride along at that velocity from then on. This "terminal velocity" is what appears in the current density formula.

The cause of this effectively viscous motion is (roughly) the many collisions that the charge carriers experience as they flow through the conductor. The charge carriers accelerate like hell for a short time, smash into something, get directed in some random direction, charge like hell again, and repeat[3]. The net *averaged* behavior is a uniform drift.

More concretely, let's look at how the momentum of charge carriers accumulate in a conductor. Consider a single charge first. Suppose at $t = 0$ a collision just occurred. Its momentum following this collision $\vec{p} = m\vec{u}_0$, where $\vec{u}_0$ is the random velocity it has following the collision. The electric field acting on this charge changes its momentum over a time interval t by imparting an impulse $\Delta\vec{p} = q\vec{E}t$. The total momentum of this particle, $\vec{p} = m\vec{u}_0 + q\vec{E}t$, thus grows linearly with time until the next collision occurs, at which point the clock is reset and we repeat.

Now consider a whole swarm of N of these charges. The *average* momentum of a member of this ensemble is

$$m\langle\vec{u}\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(m\vec{u}_i + q\vec{E}t_i\right) \ .$$

The velocity $\vec{u}_i$ is the random velocity imparted to charge i following its last collision; t_i is the time interval for that charge since that collision. If N is a large number, the first term *must be extremely close to zero.* Why? We're adding up a enormous number of randomly oriented vectors! The result of such addition is always a number that is very, very close to zero, provided N is large[4]. What remains is

$$m\langle\vec{u}\rangle = \frac{q\vec{E}}{N}\sum_{i=1}^{N} t_i \equiv q\vec{E}\tau \ ,$$

where

$$\tau \equiv \frac{1}{N}\sum_{i=1}^{N} t_i$$

is the average time between collisions. This turns out to be a property of the material; this, then, is what causes an imposed electric field to lead to uniform velocity.

[3]If you've ever driven through Los Angeles during rush hour, you've experienced the life of a charge carrier in a conductor.

[4]A careful analysis of this kind of addition shows that there is some error incurred by assuming that the vectors sum to precisely zero. However, that error is of order $1/\sqrt{N}$. As long as N is very large — which is the case for any good conductor, where N is of order Avogadro's number — these small errors are not important. Those of you who take 8.044 will learn more about this.

Notice that from this derivation, we can just read off the conductivity:

$$\begin{aligned}
\vec{J} &= nq\langle\vec{u}\rangle = nq\left(\frac{q\vec{E}\tau}{m}\right) = \sigma\vec{E} \\
\rightarrow \sigma &= \frac{nq^2\tau}{m} \,.
\end{aligned}$$

If we have multiple charge carriers, we find

$$\sigma = \sum_k \frac{n_k q_k^2 \tau_k}{m_k} \,.$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

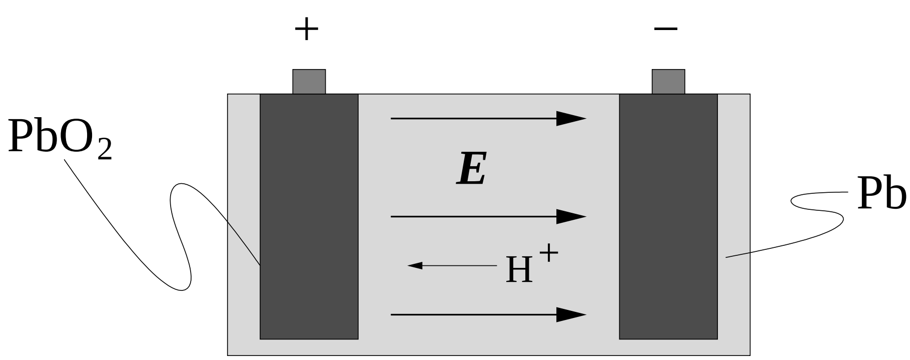
LECTURE 8:
EMF, CIRCUITS, KIRCHHOFF'S RULES

8.1 EMF: Electromotive force

The workings of our technological civilization are largely based on the idea that we use electric currents as a means to get devices to do work for us. In this lecture, we will begin to understand at a fundamental level how this works.

In order for a current to flow, we must create a potential difference over some conductive material; and, we must have continuous source of charges that can flow through our conductor. Such a source, with a "built-in" potential difference is called a source of "electromotive force" (a rather misleading name, since it isn't a force). This is usually abbreviated EMF, and is often denoted $\mathcal{E}$.

The simplest example of an EMF source is a battery. A battery is a device that maintains a separation of charges between two terminals. Current can flow inside the battery from one terminal to another via an electrochemical reaction. One example, used in many car batteries, is based on the electrochemistry of lead. Two terminals, one of lead oxide, another of porous lead, are immersed in sulphuric acid.



When immersed in the acid, it is energetically favorable for the porous lead terminal to provide free electrons, producing lead sulfate and free hydrogen ions in the solution:

$$\mathrm{Pb} + \mathrm{HSO_4^-} \rightarrow \mathrm{PbSO_4} + \mathrm{H^+} + 2\mathrm{e^-}$$

At the lead oxide electrode, a reaction which absorbs free electrons and free hydrogen ions is energetically favorable:

$$\mathrm{PbO_2} + 3\mathrm{H^+} + \mathrm{HSO_4^-} + 2\mathrm{e^-} \rightarrow \mathrm{PbSO_4} + 2\mathrm{H_2O}$$

If it is possible for both electrons and $\mathrm{H^+}$ ions to travel from one terminal to the other, then the net chemical reaction

$$\mathrm{Pb} + \mathrm{PbO_2} + 2\mathrm{H_2SO_4} \rightarrow 2\mathrm{PbSO_4} + 2\mathrm{H_2O}$$

can proceed.

When the terminals of the battery shown above are not connected, there is no way for electrons to get from one terminal to another. An electric field inside the battery builds up, pointing from the + terminal to the − terminal. This field opposes the motion of H^+ ions — they cannot cross to the + terminal, and the reaction stops. When the terminals are connected by a conductor, on the other hand, electrons freely flow to the + terminal. The electric field is reduced, the H^+ ions can now easily move across, and the reaction runs happily.

The EMF of the battery is just the potential of the + terminal with respect to the − terminal:

$$\mathcal{E} = -\int_{-\text{ term}}^{+\text{ term}} \vec{E} \cdot d\vec{s}\ .$$

This EMF is then the potential difference that is available to drive currents in an electrical circuit.

8.2 Circuits and Kirchhoff's second rule

Suppose we take a battery that provides EMF V and connect its terminals with conducting material of resistance R:

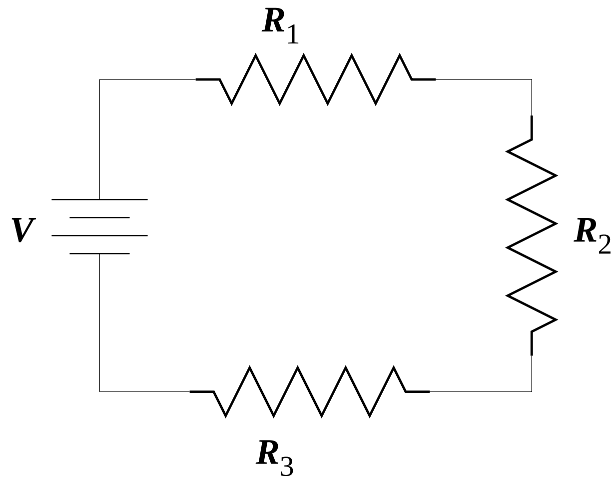
(The symbol on the left is used for a battery in a circuit; the long bar denotes the + terminal, the short bar denotes −. The symbol on the right describes a "resistor", some conducting material with known resistance R.) How much current flows in this circuit? From Ohm's law, we must have

$$V = IR \rightarrow I = V/R\ .$$

Another way to view this is to say that when a current I flows through a resistance R, there is a "voltage drop" $\Delta V = -IR$. This motivates what is known as "Kirchhoff's second[1] rule" for circuits: **Around any closed loop, the sum of the EMFs and the potential drops across circuit elements must equal zero.** This rule was first deduced by the German scientist G. Kirchhoff in the mid 1800s; really, it is just a restatement of the rule $\oint \vec{E}\cdot d\vec{s} = 0$.

[1]We'll get to Kirchhoff's first rule soon enough.

Suppose we have multiple resistors in our circuit:



The battery still supplies an EMF V to the circuit. Now, it must be shared among all three resistors. To figure out how much voltage drop occurs across each resistor, notice that the current must be the same through all of the circuit elements — if it were not the same, charge would pile up somewhere in the circuit. The voltage drop across the jth resistor is $\Delta V_j = -IR_j$; for the whole circuit, Kirchhoff's second rule tells us

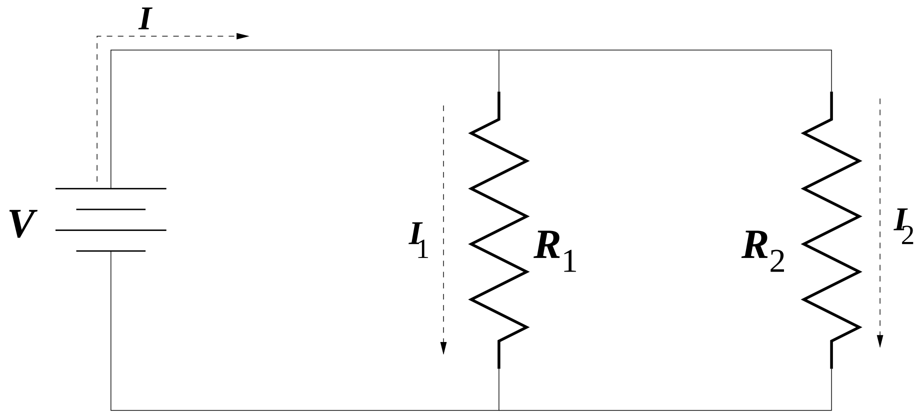
$$\begin{aligned} V &+ \sum_j \Delta V_j = 0 \\ V &- I(R_1 + R_2 + R_3) = 0 \\ \rightarrow I &= \frac{V}{R_1 + R_2 + R_3} \ . \end{aligned}$$

This is equivalent to a circuit containing a *single* resistor, with $R_{\rm eq} = R_1 + R_2 + R_3$. This logic obviously generalizes: if we have N resistors hooked up in series like this, the equivalent resistance of the circuit is given by

$$R_{\rm eq} = \sum_{i=1}^{N} R_i \ .$$

8.3 Circuits and Kirchhoff's first rule

In the previous example, the resistors were connected in *series*. Another important network topology comes from connecting the resistors in *parallel*:



The battery in this circuit supplies a total current I to all circuit elements. This current divides up into pieces I_1 and I_2. Kirchhoff #2 tells us that the sum of the EMFs and the voltage drops around *any* closed loop must be zero. A circuit network like that shown here contains three loops to which we can apply this rule: the left loop, the right loop, and the loop that goes around the perimeter of the circuit. The left loop tells us

$$V = I_1 R_1 \rightarrow I_1 = \frac{V}{R_1} \;;$$

the perimeter loop tells us

$$V = I_2 R_2 \rightarrow I_2 = \frac{V}{R_2} \;.$$

(The right loop tells us $I_1 R_1 = I_2 R_2$, which we already knew.)

To make further progress, we need to introduce Kirchhoff's first rule: **The sum of the currents entering any junction must equal the sum of the currents coming out of the junction.** This is a kind of "well, duh" rule: if it were not true, charge would pile up like mad in the circuit, which we know can't happen in the steady state. In the case of this circuit, it tells us that $I = I_1 + I_2$, so

$$I = V\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \;.$$

The resistors in parallel act like a single resistor with equivalent resistance $R_{\rm eq}$ given by

$$\frac{1}{R_{\rm eq}} = \frac{1}{R_1} + \frac{1}{R_2} \;.$$

More generally, if you have N resistors in parallel, their equivalent resistance is

$$\frac{1}{R_{\rm eq}} = \sum_{i=1}^{N} \frac{1}{R_i} \;.$$

In series, multiple resistors add, increasing the net resistance — there is only one path for the current to follow, and you are making that path "harder" by increasing its resistance. In parallel, multiple resistors add reciprocals, *decreasing* the net resistance. Rather than making the path "harder" for the current to follow, you are opening up new paths for it.

In many (though *not* all) cases, one can use the parallel and series rules to reduce a circuit to an "equivalent circuit" that is much simpler: every time you encounter resistors in series, replace them with an equivalent resistor equal to their sum; every time you encounter resistors in parallel, replace them with equivalent resistor whose reciprocal resistance equals the sum of their reciprocal resistance. In some cases, combinations of resistances that look absolutely horrible can be reduced to a single equivalent resistor; see, for example, Figure 4.16 of Purcell.

8.4 Power dissipation in circuits

As current flows in the circuit, it moves through a potential difference. Work must be done for this current to keep flowing. If the source of EMF drives charge dq through the potential difference V, the amount of work that is done is $dW = V\,dq$. If this is done in time dt, the *rate* at which work is done — i.e., the *power* that is exerted — is given by

$$P = \frac{dW}{dt} = V\,\frac{dq}{dt} = V\,I\ .$$

For a circuit element that obeys Ohm's law, this tells us that

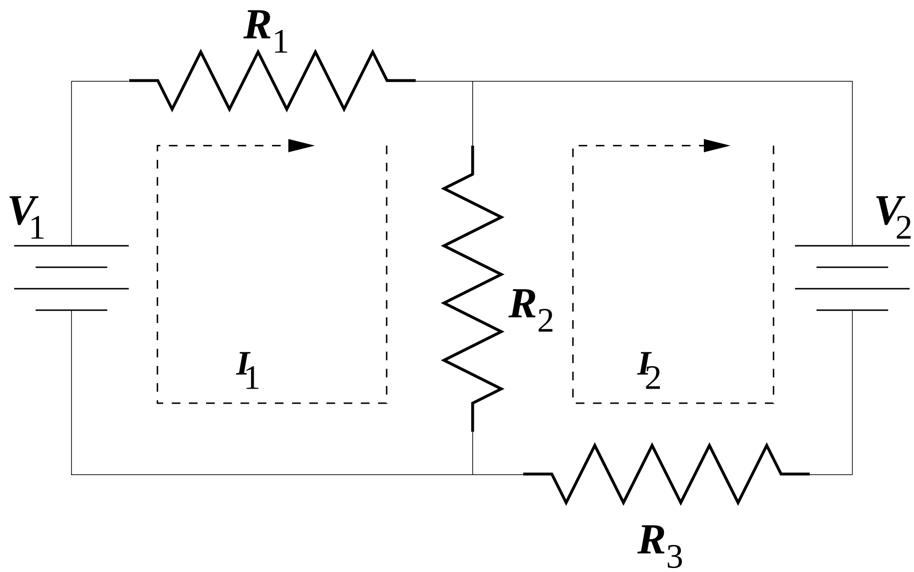
$$P = (IR)\,I = I^2 R\ .$$

This is also the power that is *dissipated* in a resistor. If you touch a circuit element and it feels hot, this is because it is dissipating energy at the rate $I^2 R$.

Units: in cgs, I is in esu/sec, and V is in esu/cm. This unit of potential is equivalent to erg/esu — recall the potential difference is work or energy per unit charge. $I\,V$ thus has units erg/sec. In SI, I is in amps, or Coulomb/sec; V is in Volts, or Joules/Coulomb. $I\,V$ thus has units Joules/sec, or Watts.

8.5 Multiple voltage source networks

How do we handle the following circuit?



The multiple sources of EMF in this circuit network make this problem somewhat more complicated than those seen so far. In particular, we cannot simply reduce the resistors into those that are in series in the circuit and those that are in parallel — the extra battery makes it hard to even define a meaningful notion of "in series" or "in parallel" in this case.

Kirchhoff's rules save us, provided we establish procedures for how to use them:

1. Regard the circuit as collection of independent closed current loops. "Independent" means that the loops may overlap, but each loop has at least one portion that does not overlap with other loops. In the example show above, there are two independent loops: the left loop and the right loop are independent (though they share one common leg); the left loop and the whole perimeter are independent; the right loop and the perimeter are independent. Once you've picked two, the third can be regarded as a subset of them.

2. Label the currents in the loops, I_1, I_2, etc, until each loop has some circulating current assigned to it. Pick the direction arbitrarily; if you guess "wrong", but do the subsequent calculation correctly, the current will just come out with a minus sign.

3. Apply Kirchhoff's #2 to each loop: traverse the loop in the direction of the current that you established. Every time you cross a resistor, you get a voltage drop $-IR$; the I to be used here is the *total* current passing through the resistor.

   Every time you cross an EMF source, it contributes $+V$ if you cross it in the "right" direction (from the $-$ terminal to the $+$). It contributes $-V$ if you cross it in the "wrong" direction ($+ \rightarrow -$).

4. Apply Kirchhoff's #1 at each junction.

Following these rules, we end up with a set of linear equations which can be reduced to find all unknown currents in the circuit (albeit sometimes requiring a hideous amount of algebra). Let's look at the circuit above as an example: Traversing the left loop, Kirchhoff #2 tells us

$$V_1 - I_1 R_1 - (I_1 - I_2) R_2 = 0 \; .$$

For the right loop, we find

$$-V_2 - I_2 R_3 - (I_2 - I_1) R_2 = 0 \; .$$

Kirchhoff's #1 is automatically satisfied, since we have $I_1 + I_2$ both entering and leaving each junction. Rearranging these equations yields

$$\begin{aligned} V_1 &= I_1 R_1 + (I_1 - I_2) R_2 \\ V_2 &= -I_2 R_3 - (I_2 - I_1) R_2 \; . \end{aligned}$$

With a little effort, we can reduce these equations to find

$$\begin{aligned} I_1 &= \frac{V_1 R_3 + (V_1 - V_2) R_2}{R_1 R_2 + R_1 R_3 + R_2 R_3} \; , \\ I_2 &= \frac{(V_1 - V_2) R_2 - V_2 R_1}{R_1 R_2 + R_1 R_3 + R_2 R_3} \; . \end{aligned}$$

Although it is somewhat beyond the scope of this course, it is worth noting that the equations we get with Kirchhoff's laws can be rewritten as a matrix equation. For the example discussed here, we would find

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} R_1 + R_2 & -R_2 \\ R_2 & -R_2 - R_3 \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \; .$$

We can use matrix algebra to solve for the currents I_1 and I_2 in terms of V_1 and V_2; the answer will of course be identical to that given above.

More generally, if there are n independent loops in the circuit, Kirchhoff's laws provide n equations that can be written in the form

$$\tilde{V} = \mathbf{R} \cdot \tilde{I} \; ,$$

where

$$\tilde{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} \; , \qquad \tilde{I} = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix} \; ,$$

and where $\mathbf{R}$ is an $n \times n$ matrix whose elements all have the units of resistance. This may sound hideous ... *but*, the solution is now relatively simple: we "simply" invert the matrix, and find the n currents from the equation

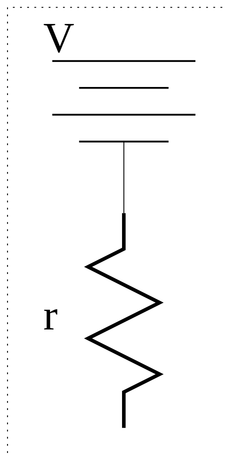
$$\tilde{I} = \mathbf{R}^{-1} \cdot \tilde{V} \; .$$

This is the only practical way to solve for the currents in circuits that have lots of loops. Doing the inversion isn't even all that bad, thanks to computerized math tools. For very large matrices ($n > 100$ or so), very fast matrix inverters exist for numerically specified matrices.

8.6 EMF and batteries revisisted: Internal resistance

One final issue of practice is often of great importance when working with batteries. When a battery is delivering current to a circuit, there is necessarily a flow of current *within the battery itself*. In the example discussed at the start of this lecture, this current comes from the motion of H^+ ions from the Pb to the PbO_2.

Because this reaction proceeds chemically, these reactions will dissipate energy; the battery gets hot, and energy which could have gone into the circuit is lost. This is described mathematically by giving the battery an *internal resistance* r. For the purpose of any calculation, the internal resistance is taken into account by simply putting a small resistor r in series with the battery. This is usually drawn something like this:



An important consequence of internal resistance is that a battery has a maximum current it can deliver: $I_{\rm max} = V/r$. Roughly speaking, its chemical reactions cannot proceed fast enough to deliver more than this current. When this happens, a power $I^2_{\rm max}r$ is dissipated in the battery, which gets hot in a hurry. A good way to prove to yourself the reality of internal resistance is to carry a nine-volt battery in a pocket full of keys and change. The pieces of metal short out the battery's terminals, it gets hot, you scream in pain and hop around like a lunatic[2].

[2]Yes, I speak from personal experience.

Scott Hughes *8 March 2005*

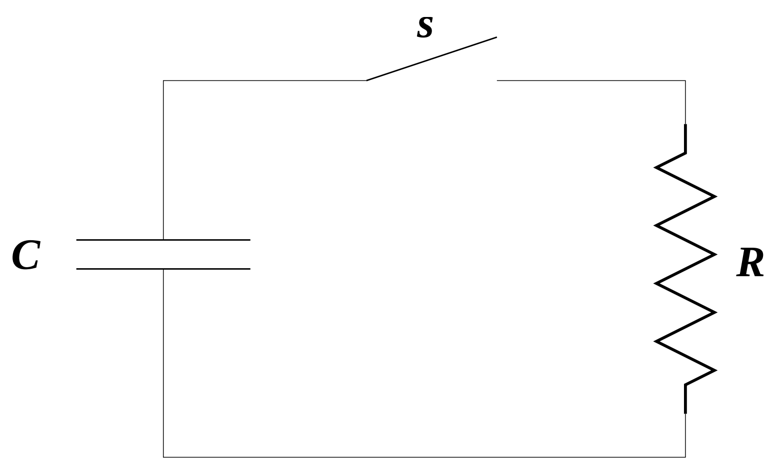
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 9:
VARIABLE CURRENTS; THÉVENIN EQUIVALENCE

9.1 Variable currents 1: Discharging a capacitor

Up til now, everything we have done has assumed that things are in steady state: all fields are constant; charges are either nailed in place or else are flowing uniformly. In reality, such a situation tends to be the exception rather than the rule. We now start to think about things that vary in time.

Suppose we have a capacitor C charged up until the potential difference between its plates is V_0; the charge separation is $Q_0 = CV_0$. We put this capacitor into a circuit with a resistor R and a switch s:



What happens when the switch is closed?

Before thinking about this with equations, let's think about what happens here physically. The instant that the switch is closed, there is a potential difference of V_0 across the resistor. This drives a current to flow. Now, this current can only come from the charge separation on the plates of the capacitor: the excess charges on one plate flow off and neutralize the deficit of charges on the other plate. The flow of current thus serves to reduce the amount of charge on the capacitor; by $Q = CV$, this must reduce the voltage across the capacitor. The potential difference which drives currents thus becomes smaller, and so the current flow should reduce. We expect to see a flow of current that starts out big and gradually drops off.

To substantiate this, turn to Kirchhoff's laws: at any moment, the capacitor supplies an EMF $V = Q/C$. As the current flows, there is a voltage drop $-IR$ across the resistor:

$$\frac{Q}{C} - IR = 0\ .$$

This is true, but not very helpful — we need to connect the charge on the capacitor Q to the flow of current I. Thinking about it a second, we see that we must have

$$I = -\frac{dQ}{dt} \; .$$

Why the minus sign? The capacitor is *losing* charge; more accurately, the charge separation is being reduced. In this situation, a large, positive current reflects a large reduction in the capacitor's charge separation.

Putting this into Kirchhoff, we end up with a first order differential equation:

$$\frac{Q}{C} + R\frac{dQ}{dt} = 0 \; .$$

To solve it, rearrange this in a slightly funny way:

$$\frac{dQ}{Q} = -\frac{dt}{RC} \; .$$

Then integrate both sides. We use the integral to enforce the *boundary conditions*: initially ($t = 0$), the charge separation is Q_0. At some later time t, it is a value $Q(t)$. Our goal is to find this $Q(t)$:

$$\int_{Q=Q_0}^{Q=Q(t)} \frac{dQ}{Q} = -\int_{t=0}^{t} \frac{dt}{RC}$$

$$\rightarrow \ln\left[\frac{Q(t)}{Q_0}\right] = -\frac{t}{RC} \; .$$

Taking the exponential of both sides gives the solution:

$$Q(t) = Q_0 e^{-t/RC} \; .$$

The charge decays exponentially. After every time interval of RC, the charge has fallen by a factor of $1/e$ relative to its value at the start of the interval.

Sanity check: does RC make sense as a time? Let's check its units:

- CGS: $R \times C$ has units (sec/cm) $\times$ (cm) = seconds.
- SI: $R \times C$ is (ohms) $\times$ (farads) = (volts/amps) $\times$ (coulombs/volts) = coulombs/amps = seconds.

So it does make sense.

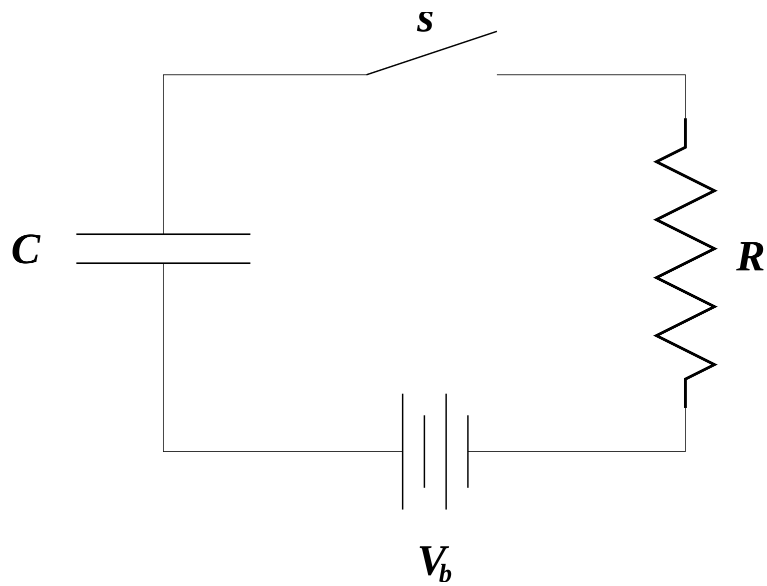
Let's look at the current $I(t)$:

$$I(t) = -\frac{dQ}{dt} = -Q_0\frac{d}{dt}e^{-t/RC} = \frac{Q_0}{RC}e^{-t/RC} \; .$$

Just like the charge, the current decays exponentially. In particular, it starts out at some big value, and then falls away — just as our physical intuition told us it would.

9.2 Variable currents 2: Charging a capacitor

Let's consider a different kind of circuit. The capacitor begins, at $t = 0$, with *no* charge; but, the circuit now contains a battery:



Now, what happens when the switch is closed?

Let's again think about this physically before doing the math. When the switch is first closed, the EMF of the battery can very easily drive a current, so we expect the current to jump up, and charge to begin accumulating on the capacitor. As this charge accumulates, an $\vec{E}$ field will build up in a direction that *opposes* the flow of current! Thus, we expect that, as time goes on, the flow of current will decrease. Eventually it will stop, when there is just enough charge on the capacitor to totally oppose the battery.

Let's check this. For Kirchhoff, we have an EMF V_b from the battery, and two voltage drops: $-Q/C$ from the capacitor, and $-IR$ across the resistor:

$$V_b - \frac{Q}{C} - IR = 0 \; .$$

We again need to relate the current I and the capacitor's charge Q. Thinking about it for a second, we must have

$$I = +\frac{dQ}{dt} \; .$$

Why a plus sign? In this case we are *adding* charge to the capacitor (more accurately, *increasing* the charge separation). Large current means a large increase in the charge separation. When doing a problem involving a circuit like this, it is important to stop and think carefully about how the capacitor's charge relates to the charge flowing onto it or off of it.

We end up with the following differential equation:

$$V_b - \frac{Q}{C} - R\frac{dQ}{dt} = 0 \; .$$

Rearrange:

$$\frac{dQ}{dt} = -\frac{Q - CV_b}{RC}$$

or

$$\frac{dQ}{Q - CV_b} = -\frac{dt}{RC} \ .$$

Integrate: we require $Q(t = 0) = 0$, so we find

$$\ln\left[\frac{CV_b - Q(t)}{CV_b}\right] = -\frac{t}{RC} \ .$$

Exponentiating both sides and solving for $Q(t)$, we find

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

This charge starts off at zero, builds up quickly at first, then starts building up more and more slowly. It asymptotically levels off at $Q = CV_b$ as $t \to \infty$. Let's look at the current:

$$\begin{aligned} I(t) &= CV_b\frac{d}{dt}\left(1 - e^{-t/RC}\right) \\ &= \frac{V_b}{R}e^{-t/RC} \ . \end{aligned}$$

As we intuitively guessed, the current starts large, but drops off as time passes and the capacitor's electric field impedes it.

Notice that the voltage across the capacitor

$$V_C(t) = \frac{Q(t)}{C} = V_b\left(1 - e^{-t/RC}\right)$$

and the voltage across the resistor

$$V_R(t) = I(t)R = V_b e^{-t/RC}$$

sum to give a constant,

$$V_C(t) + V_R(t) = V_b\left(1 - e^{-t/RC}\right) + V_b e^{-t/RC} = V_b \ ,$$

the EMF of the battery. If you think about Kirchhoff's rules for a moment, this should make a lot of sense!

9.3 Charging and discharging revisited

Suppose you're taking a test and you don't want to slog through setting up a bunch of integrals and running the risk of botching the details because of a silly mistake. Provided you understand *physically* what is going on in the above discussions, you can write down the rules for charge on a capacitor without doing any math. You only need to know three things:

1. The time variation has to look like $e^{-t/RC}$.
2. You need the initial ($t = 0$) charge on the capacitor, and
3. You need the final ($t \to \infty$) charge on the capacitor.

Then, you just need to figure out to hook everything together so that you get a formula that obeys the correct boundary conditions. For example, if a problem involves discharging a capacitor, you know that you start out with some initial value Q_0, and that it must fall towards zero as time passes. The only formula that obeys these conditions and has the correct time variation is

$$Q(t) = Q_0 e^{-t/RC} ,$$

just what we derived carefully before. If it involves charging up a capacitor, you want a formula that has $Q = 0$ at $t = 0$, and that levels off at CV_b as $t \to \infty$. Using $1 - e^{-t/RC}$ as our "time variation" piece of the solution gets the $t = 0$ behavior right; multiplying by CV_b ensures that we level off at the right value at large t. We end up with

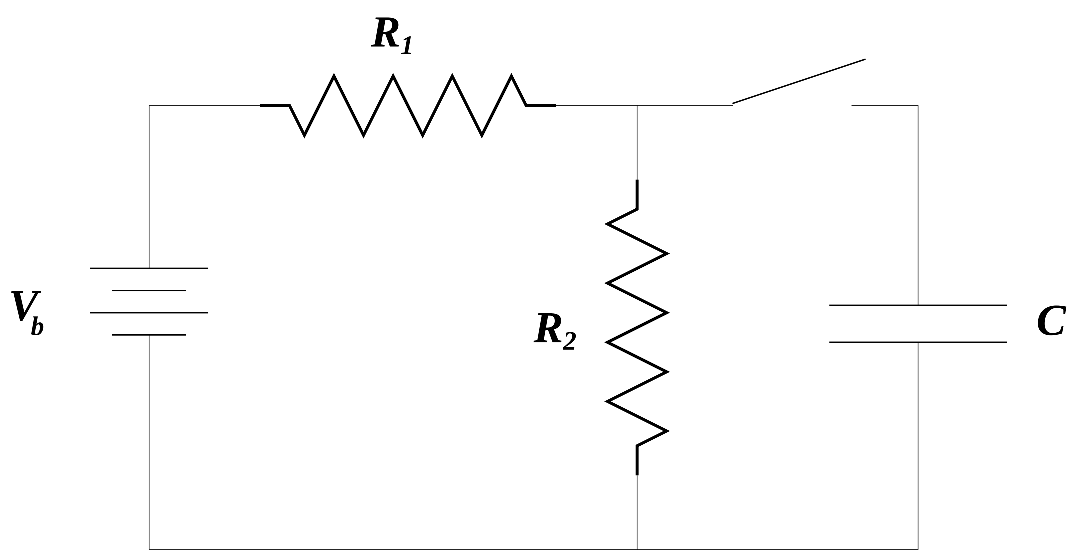
$$Q(t) = CV_b \left(1 - e^{-t/RC}\right) .$$

Again, just what we derived carefully before.

If you can remember these physical reasons why the charge behaves as it does, you can save yourself some pain later on.

9.4 Thévenin equivalence

The RC circuits we've looked at so far are quite simple — there's only one resistor, and it's hooked up to the capacitor and/or battery in series. How do we handle a more complicated circuit? Consider the following:



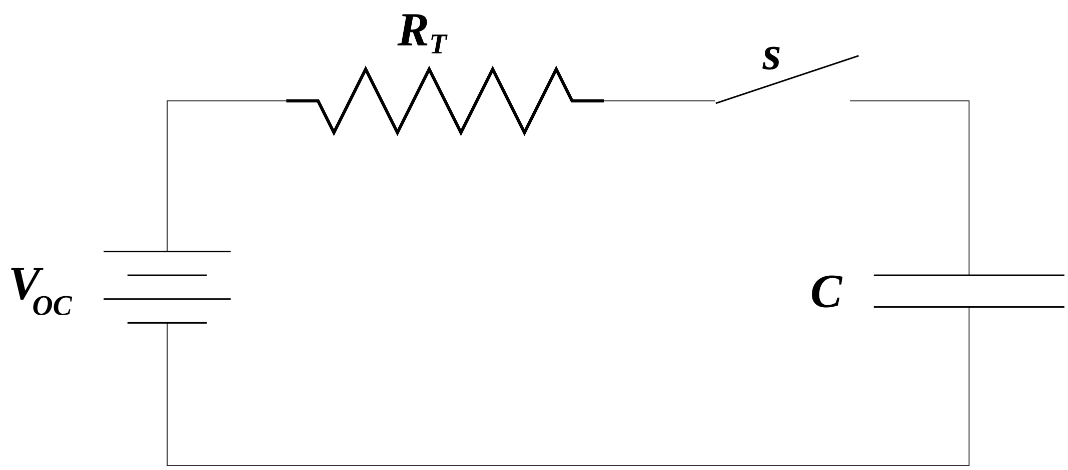
How does the charge on the capacitor evolve with time?

This problem can be tackled quite simply by using Kirchhoff's laws: we have multiple loops, so we go around them and write down equations balancing the voltage drops and EMFs; we make sure the currents balance at the junctions; math happens; we find $Q(t)$. As an exercise, this is straightforward, but slightly tedious.

There is in fact a very cute way to approach this, based on a theorem proved by Léon Charles Thévenin[1]:

Thévenin's Theorem: Any combination of batteries and resistances with two terminals can be replaced by a single voltage source $V_{\rm OC}$ and a single series resistor R_T.

For our problem, this means that we can redraw the circuit as



The answer is now obvious! We just steal results from before, and write down

$$Q(t) = CV_{\rm OC}\left(1 - e^{-t/R_T C}\right) \ .$$

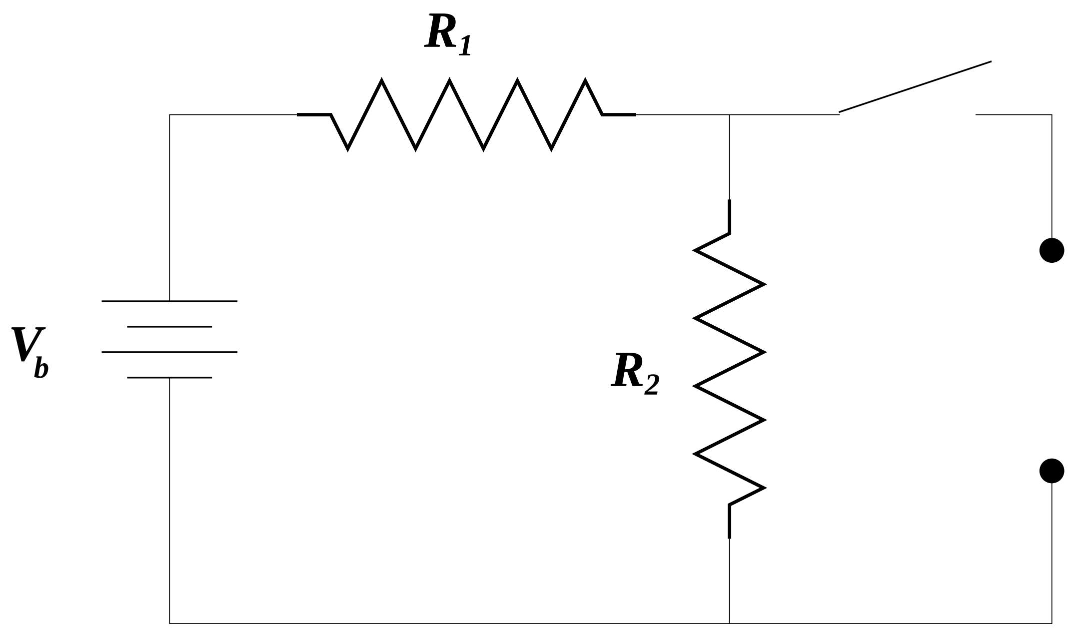
We just need to know how to work out $V_{\rm OC}$ and R_T.

[1]I tried to find biographical information about this guy and totally failed. Fortunately, an anonymous Spring '04 8.022 student was more persistant than me, and found a web bio:

`http://www.eleceng.adelaide.edu.au/famous.html#thevenin`

He was a French telegraph engineer, and hence very well versed in practical aspects of electrical circuits. Interestingly, "his" theorem was actually first derived by Hermann Von Helmholtz.

To do this, we take the capacitor out of the circuit and consider what remains as a pair of terminals:



If we short circuit these terminals with a wire, a current I_{SC} — the "short circuit" current — will flow. On the other hand, if we leave it open, there will be a potential difference — the "open circuit" voltage, V_{OC} — between the terminals (equivalent to the potential drop over the resistor R_2).

Let's calculate these quantities. Shorting the terminals out is equivalent to putting a 0 Ohm resistor in parallel with R_2. The current that flows is what you would get if R_2 were not even present:

$$I_{\mathrm{SC}} = \frac{V_b}{R_1} \; .$$

On the other hand, if we leave the terminals open, the current that flows through the circuit is $I = V_b/(R_1 + R_2)$. The open circuit voltage is thus given by

$$V_{\mathrm{OC}} = V_{R_2} = IR_2 = \frac{V_b R_2}{R_1 + R_2} \; .$$

The *Thévenin equivalent resistance* is defined as the ratio of these quantities:

$$R_T \equiv \frac{V_{\mathrm{OC}}}{I_{\mathrm{SC}}} \; .$$

For this particular circuit, it takes the value

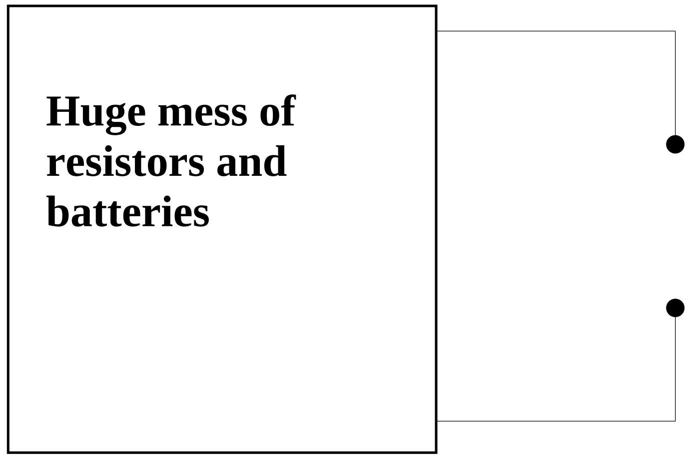
$$\begin{aligned} R_T &= \frac{V_b R_2}{R_1 + R_2} \times \frac{R_1}{V_b} \\ &= \frac{R_1 R_2}{R_1 + R_2} \; . \end{aligned}$$

The solution for the charge on the capacitor is thus finally given by

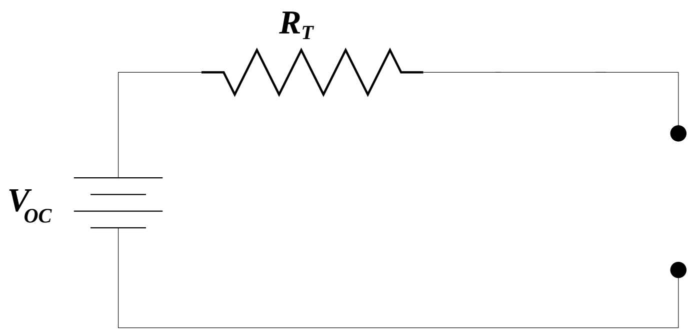
$$Q(t) = CV_b \left(\frac{R_2}{R_1 + R_2} \right) \left[1 - \exp(-t(R_1 + R_2)/CR_1R_2) \right] \; .$$

9.5 Thévenin in general

Thévenin's theorem is much more general than this simple example. In general, it means that you can be given any grungy mess of resistors and batteries and reduce it to a single battery with EMF V_{OC} and a single resistor R_T: anything like this



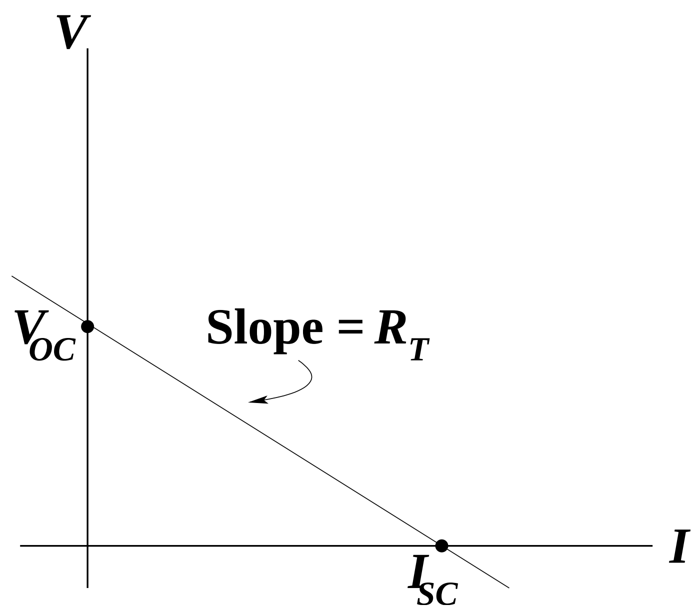
can be reduced to this



If you know what the big mess inside the "black box" is made out of, then you can figure out V_{OC} and R_T as we did for the example above.

In some situations, you *don't* know. For example, you might have a piece of lab equipment that you need to use for some experiment, and you need to know how to treat it in a circuit. In that case, you measure V_{OC}; you short out the terminals and measure I_{SC}; and then you know $R_T = V_{\mathrm{OC}}/I_{\mathrm{SC}}$.

Thévenin's theorem works only when all the elements inside the box obey Ohm's law. This means that the relationship between current and voltage for every element is *linear*. This in turn means that the relationship between current and voltage for any *combination* of circuit elements in the box must be linear. (Why? Add a bunch of lines together: you get a line!) The procedure for determining the Thévenin equivalence of a circuit simply assumes that you have some linear relationship between the voltage and current at the terminals. V_{OC} tells us where this relationship crosses the voltage axis; I_{SC} tells us where it crosses the current axis. R_T is the slope. (Negative slope, really, since the line rises in the "wrong" direction; same difference.)


V
V_{OC}
Slope = R_T
I
I_{SC}

Scott Hughes *10 March 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2004

LECTURE 10:
MAGNETIC FORCE; MAGNETIC FIELDS; AMPERE'S LAW

10.1 The Lorentz force law

Until now, we have been concerned with electrostatics — the forces generated by and acting upon charges at rest. We now begin to consider how things change when charges are in motion[1].

A simple apparatus demonstrates that something wierd happens when charges are in motion: If we run currents next to one another in parallel, we find that they are *attracted* when the currents run in the same direction; they are *repulsed* when the currents run in opposite directions. This is despite the fact the wires are *completely neutral*: if we put a stationary test charge near the wires, it feels no force.

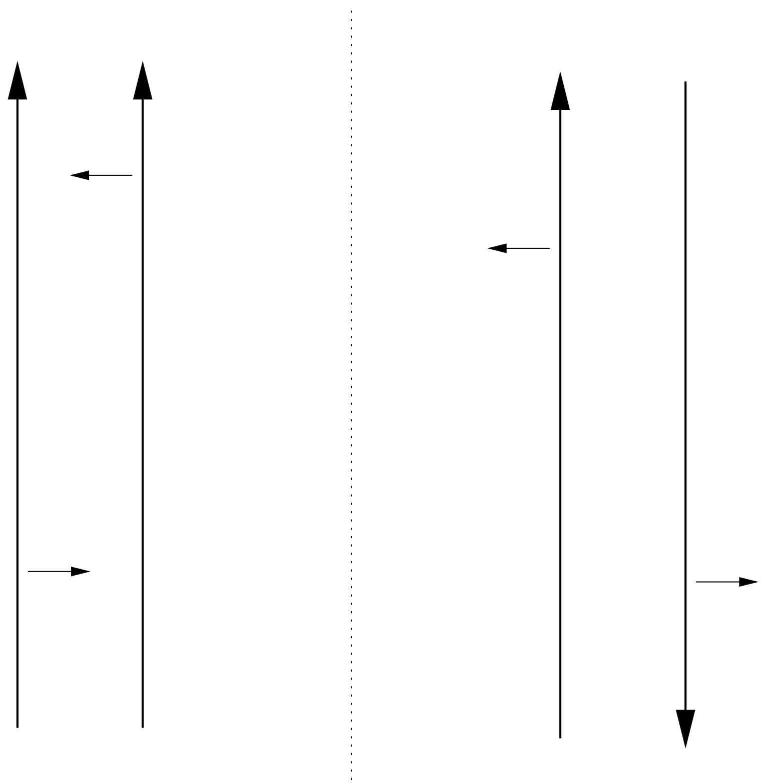


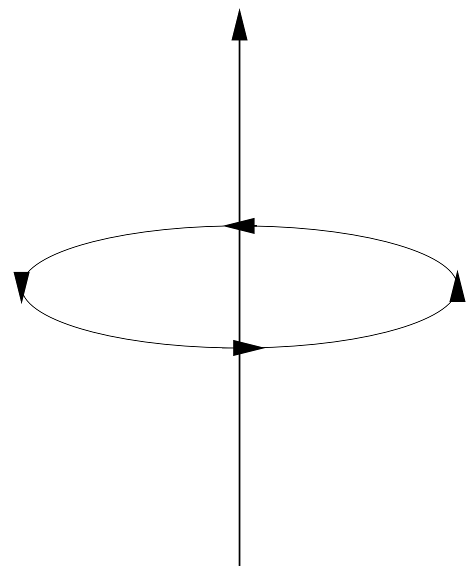
Figure 1: Left: parallel currents attract. Right: Anti-parallel currents repel.

Furthermore, experiments show that the force is proportional to the currents — double the current in *one* of the wires, and you double the force. Double the current in both wires, and you quadruple the force.

[1]We will deviate a bit from Purcell's approach at this point. In particular, we will defer our discussion of special relativity til next lecture.

This all indicates a force that is proportional to the velocity of a moving charge; and, that points in a direction *perpendicular* to the velocity. These conditions are screaming for a force that depends on a cross product.

What we say is that some kind of field $\vec{B}$ — the "magnetic field" — arises from the current. (We'll talk about this in detail very soon; for the time being, just accept this.) The direction of this field is kind of odd: it wraps around the current in a circular fashion, with a direction that is defined by the right-hand rule: We point our right thumb in the direction of the current, and our fingers curl in the same sense as the magnetic field.



With this sense of the magnetic field defined, the force that arises when a charge moves through this field is given by

$$\vec{F} = q\frac{\vec{v}}{c} \times \vec{B} ,$$

where c is the speed of light. The appearance of c in this force law is a hint that special relativity plays an important role in these discussions.

If we have both electric and magnetic fields, the *total* force that acts on a charge is of course given by

$$\vec{F} = q\left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B}\right) .$$

This combined force law is known as the Lorentz force.

10.1.1 Units

The magnetic force law we've given is of course in cgs units, in keeping with Purcell's system. The magnetic force equation itself takes a slightly different form in SI units: we do not include the factor of $1/c$, instead writing the force

$$\vec{F} = q\vec{v} \times \vec{B} .$$

This is a very important difference! It makes comparing magnetic effects between SI and cgs units slightly nasty.

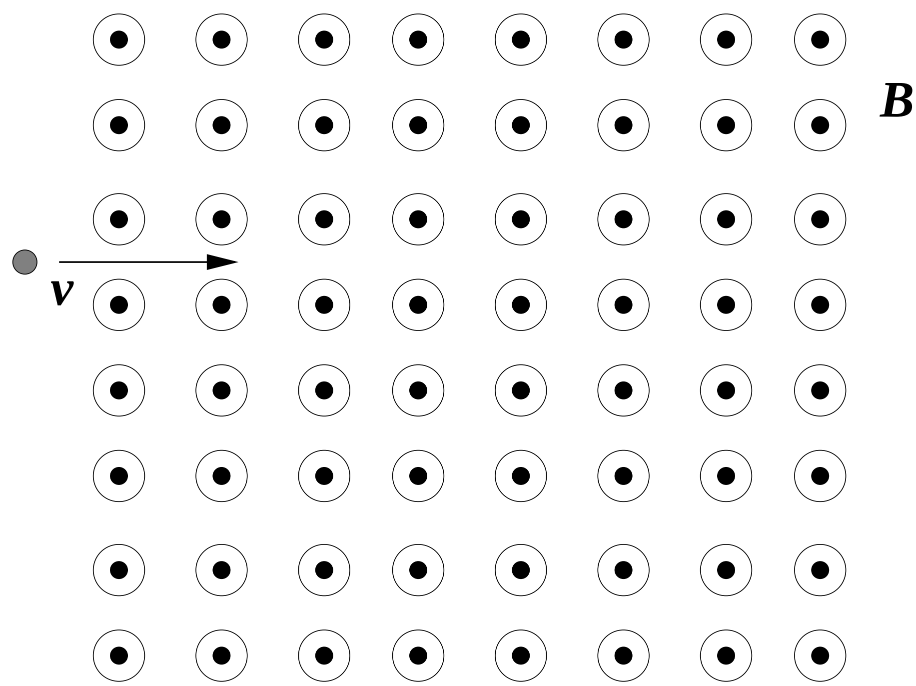
Notice that, in cgs units, the magnetic field has the same overall dimension as the electric field: $\vec{v}$ and c are in the same units, so $\vec{B}$ must be force/charge. For historical reasons, this combination is given a special name: 1 dyne/esu equals 1 Gauss (1 G) when the force in question is magnetic. (There is no special name for this combination when the force is electric.)

In SI units, the magnetic field does *not* have the same dimension as the electric field: $\vec{B}$ must be force/(velocity $\times$ charge). The SI unit of magnetic field is called the Tesla (T): the Tesla equals a Newton/(coulomb $\times$ meter/sec).

To convert: $1\,\mathrm{T} = 10^4\,\mathrm{G}$.

10.2 Consequences of magnetic force

Suppose I shoot a charge into a region filled with a uniform magnetic field:



The magnetic field $\vec{B}$ points out of the page; the velocity $\vec{v}$ initially points to the right. What motion results from the magnetic force?

At every instant, the magnetic force points perpendicular to the charge's velocity — exactly the force needed to cause circular motion. It is easy to find the radius of this motion: if the particle has charge q and mass m, then

$$\begin{aligned} F_{\mathrm{mag}} &= F_{\mathrm{centripetal}} \\ \frac{qvB}{c} &= \frac{mv^2}{R} \\ R &= \frac{mvc}{qB}\,. \end{aligned}$$

If the charge q is positive, the particle's trajectory veers to the right, vice versa if its negative. This kind of qualitative behavior — bending the motion of charges along a curve — is typical of magnetic forces.

Notice that magnetic forces do no work on moving charges: if we imagine the charge moves for a time dt, the work that is done is

$$\begin{aligned} dW &= \vec{F} \cdot d\vec{s} = \vec{F} \cdot \vec{v}\, dt \\ &= q\left(\frac{\vec{v}}{c} \times \vec{B}\right) \cdot \vec{v}\, dt \\ &= 0\,. \end{aligned}$$

The zero follows from the fact that $\vec{v} \times \vec{B}$ is perpendicular to $\vec{v}$.

10.3 Force on a current

Since a current consists of a stream of freely moving charges, a magnetic field will exert a force upon any flowing current. We can work out this force from the general magnetic force law.

Consider a current I that flows down a wire. This current consists of some linear density of freely flowing charges, λ, moving with velocity $\vec{v}$. (The direction of the charges' motion is defined by the wire: they are constrained by the wire's geometry to flow in the direction it points.) Look at a little differential length dl of this wire (a vector, since the wire defines the direction of current flow).

The amount of charge contained in this differential length is $dq = \lambda\, dl$. The differential of force exerted on this piece of the wire is then

$$d\vec{F} = (\lambda\, dl)\frac{\vec{v}}{c} \times \vec{B}\,.$$

There are two equivalent ways to rewrite this in terms of the current. First, because the current is effectively a vector by virtue of the velocity of its constituent charges, we put $\vec{I} = \lambda\vec{v}$ and find

$$d\vec{F} = dl\,\frac{\vec{I}}{c} \times \vec{B}\,.$$

Second, we can take the current to be a scalar, and use the geometry of the wire to define the vector:

$$d\vec{F} = \frac{I}{c} d\vec{l} \times \vec{B}\,.$$

These two formulas are completely equivalent to one another. Let's focus on the second version. The total force is given by integrating:

$$\vec{F} = \frac{I}{c}\int d\vec{l} \times \vec{B}\,.$$

If we have a long, straight wire whose length is L and is oriented in the $\hat{n}$ direction, we find

$$\vec{F} = \frac{IL}{c}\hat{n} \times \vec{B}\,.$$

This formula is often written in terms of the force per unit length: $\vec{F}/L = (I/c)\hat{n} \times \vec{B}$.

10.4 Ampere's law

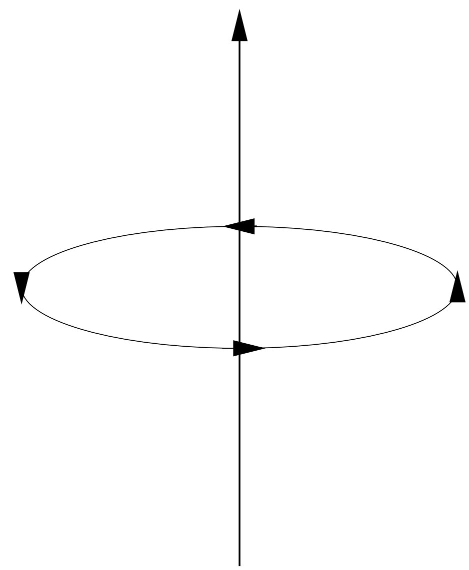
We've talked about the force that a magnetic field exerts on charges and current; but, we have not yet said anything about where this field comes from. I will now give, without *any* proof or motivation, a few key results that allow us to determine the magnetic field in many situations.

The main result we need is *Ampere's law*:

$$\oint_C \vec{B} \cdot d\vec{s} = \frac{4\pi}{c} I_{\rm encl} \ .$$

In words, if we take the line integral of the magnetic field around a closed path, it equals $4\pi/c$ times the current enclosed by the path.

Ampere's law plays a role for magnetic fields that is similar to that played by Gauss's law for electric fields. In particular, we can use it to calculate the magnetic field in situations that are sufficiently symmetric. An important example is the magnetic field of a long, straight wire: In this situation, the magnetic field must be constant on any circular path around the



wire. The amount of current enclosed by this path is just I, the current flowing in the wire:

$$\oint \vec{B} \cdot d\vec{s} = B(r) 2\pi r = \frac{4\pi}{c} I$$

$$\rightarrow B(r) = \frac{2I}{cr} \ .$$

The magnetic field from a current thus falls off as $1/r$. Recall that we saw a similar $1/r$ law not so long ago — the electric field of a long line charge also falls off as $1/r$. As we'll see fairly soon, this is not a coincidence.

The direction of this field is in a "circulational" sense — the $\vec{B}$ field winds around the wire according to the right-hand rule[2]. This direction is often written $\hat{\phi}$, the direction of

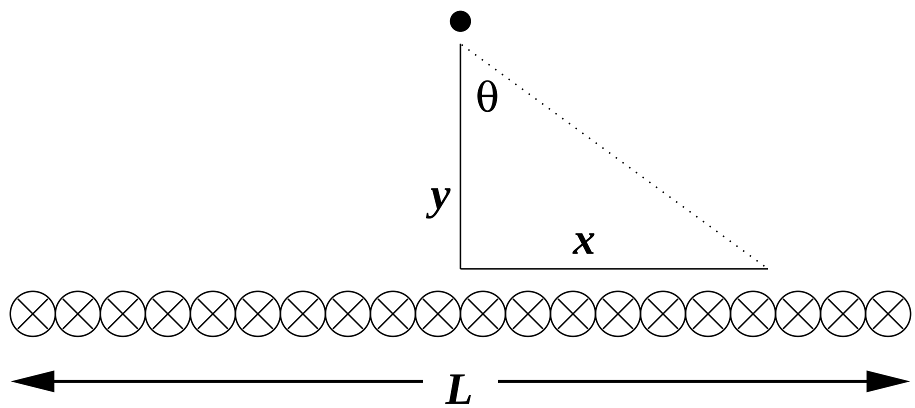
[2]In principle, we could have defined it using a "left-hand rule". This would give a fully consistent description of physics provided we switched the order of all cross products (which is identical to switching the sign of all cross products).

increasing polar angle ϕ. The full vector magnetic field is thus written

$$\vec{B} = \frac{2I}{cr}\hat{\phi} \, .$$

10.4.1 Field of a plane of current

The magnetic field of the long wire can be used to derive one more important result. Suppose we take a whole bunch of wires and lay them next to each other:



The current in each wire is taken to go into the page. Suppose that the total amount of current flowing in *all* of the wires is I, so that the current *per unit length* is $K = I/L$. What is the magnetic field at a distance y above the center of the plane?

This is fairly simple to work out using superposition. First, from the symmetry, you should be able to see that only the horizontal component of the magnetic field (pointing to the right) will survive. For the vertical components, there will be equal and opposite contributions from wires left and right of the center. To sum what's left, we set up an integral:

$$\vec{B} = \frac{2}{c}\hat{x}\int_{-L/2}^{L/2} \frac{(I/L)dx}{\sqrt{x^2+y^2}}\cos\theta \, .$$

In the numerator under the integral, we are using $(I/L)dx$ as the current carried by a "wire" of width dx. Doing the trigonometry, we replace $\cos\theta$ with something a little more useful:

$$\begin{aligned}\vec{B} &= \frac{2}{c}\hat{x}\int_{-L/2}^{L/2} \frac{(I/L)dx}{\sqrt{x^2+y^2}}\frac{y}{\sqrt{x^2+y^2}} \\ &= \frac{2Ky}{c}\hat{x}\int_{-L/2}^{L/2}\frac{dx}{x^2+y^2} \, .\end{aligned}$$

This integral is doable, but not particularly pretty (you end up with a mess involving arctangents). A more tractable form is obtaining by taking the limit of $L \to \infty$: using

$$\int_{-\infty}^{\infty}\frac{dx}{x^2+y^2} = \frac{\pi}{|y|} \, ,$$

we find

$$\vec{B} = \hat{x}\frac{2\pi K y}{c|y|}$$

$$
\begin{aligned}
&= +\hat{x}\frac{2\pi K}{c} \qquad y > 0 \quad \text{(above the plane)} \\
&= -\hat{x}\frac{2\pi K}{c} \qquad y < 0 \quad \text{(below the plane)}
\end{aligned}
$$

The most important thing to note here is the change as we cross the sheet of current:

$$|\Delta\vec{B}| = \frac{4\pi K}{c} \ .$$

Does this remind you of anything? It should! When we cross a sheet of *charge* we have

$$|\Delta\vec{E}| = 4\pi\sigma \ .$$

The sheet of current plays a role in magnetic fields very similar to that played by the sheet of charge for electric fields.

10.4.2 Force between two wires

Combining the result for the magnetic field from a wire with current I_1 with the force per unit length upon a long wire with current I_2 tells us the force per unit length that arises between two wires:

$$\frac{|\vec{F}|}{L} = \frac{2I_1I_2}{c^2r} \ .$$

Using right-hand rule, you should be able to convince yourself quite easily that this force is attractive when the currents flow in the same direction, and is repulsive when they flow in opposite directions.

10.4.3 SI units

In SI units, Ampere's law takes the form

$$\oint_C \vec{B}\cdot d\vec{S} = \mu_0 I_{\rm encl}$$

where the constant $\mu_0 = 4\pi \times 10^{-7}$ Newtons/amp^2 is called the "magnetic permeability of free space". To convert any cgs formula for magnetic field to SI, multiply by $\mu_0 \times (c/4\pi)$. For example, the magnetic field of a wire becomes

$$\vec{B} = \frac{\mu_0 I}{2\pi r}\hat{\phi} \ .$$

The force between two wires becomes

$$\frac{|\vec{F}|}{L} = \frac{\mu_0 I_1 I_2}{2\pi r} \ .$$

If you try to reproduce this force formula, remember that the magnetic force in SI units does not have the factor $1/c$.

10.5 Divergence of the $\vec{B}$ field

Let's take the divergence of straight wire's magnetic field using Cartesian coordinates. We put $r = \sqrt{x^2 + y^2}$. With a little trigonometry, you should be able to convince yourself that

$$\begin{aligned} \hat{\phi} &= \hat{y}\cos\phi - \hat{x}\sin\phi \\ &= \frac{x\hat{y}}{\sqrt{x^2+y^2}} - \frac{y\hat{x}}{\sqrt{x^2+y^2}} , \end{aligned}$$

so

$$\vec{B} = \frac{2I}{c}\left[\frac{x\hat{y}}{x^2+y^2} - \frac{y\hat{x}}{x^2+y^2}\right] .$$

The divergence of this field is

$$\begin{aligned} \vec{\nabla}\cdot\vec{B} &= \frac{2I}{c}\left[\frac{2yx}{(x^2+y^2)^2} - \frac{2xy}{(x^2+y^2)^2}\right] \\ &= 0 . \end{aligned}$$

We could have guessed this without doing any calculation: if we make any small box, there will be just as many field lines entering it as leaving.

Although we have only done this in detail for this very special case, it turns out this result holds for magnetic fields in general:

$$\boxed{\boxed{\vec{\nabla}\cdot\vec{B} = 0}}$$

Recall that we found $\vec{\nabla}\cdot\vec{E} = 4\pi\rho$ — the divergence of the electric field told us about the density of electric charge. The result $\vec{\nabla}\cdot\vec{B} = 0$ thus tells us that there is *no such thing as magnetic charge*.

This is actually not a foregone conclusion: there are reasons to believe that very small amounts of magnetic charge may exist in the universe, created by processes in the big bang. If any such charge exists, it would create a "Coulomb-like" magnetic field, with a form just like the electric field of a point charge. No conclusive evidence for these monopoles has ever been found; but, absence of evidence is not evidence of absence.

10.6 What *is* the magnetic field???

This "magnetic field" is, so far, just a construct that may seem like I've pulled out of the air. I haven't pulled it out of the air just for kicks — observations and measurements demonstrate that there is an additional field that only acts on moving charges. But what exactly is this field? Why should there exist some field that only acts on *moving* charges?

The answer is to be found in special relativity. The defining postulate of special relativity essentially tells us that physics must be *consistent* in every "frame of reference". Frames of reference are defined by observers moving, with respect to each other, at different velocities.

Consider, for example, a long wire in some laboratory that carries a current I. In this "lab frame", the wire generates a magnetic field. Suppose that a charge moves with velocity $\vec{v}$ parallel to this wire. The magnetic field of the wire leads to an attractive force between the charge and the wire.

Suppose we now examine this situation from the point of view of the charge (the "charge frame"). From the charge's point of view, it is sitting perfectly still. If it is sitting still, **there can be no magnetic force!** We appear to have a problem: in the "lab frame", there is an attractive magnetic force. In the "charge frame", there can't possibly be an attractive magnetic force. But for physics to be consistent in both frames of reference, there must be *some* attractive force in the charge frame. What is it???

There's only thing it can be: in the charge's frame of reference, **there must be an attractive ELECTRIC field.** In other words, what looks like a pure magnetic field in one frame of reference looks (at least in part) like an electric field in another frame of reference. To understand how this happens, we must begin to understand special relativity. This is our next topic.

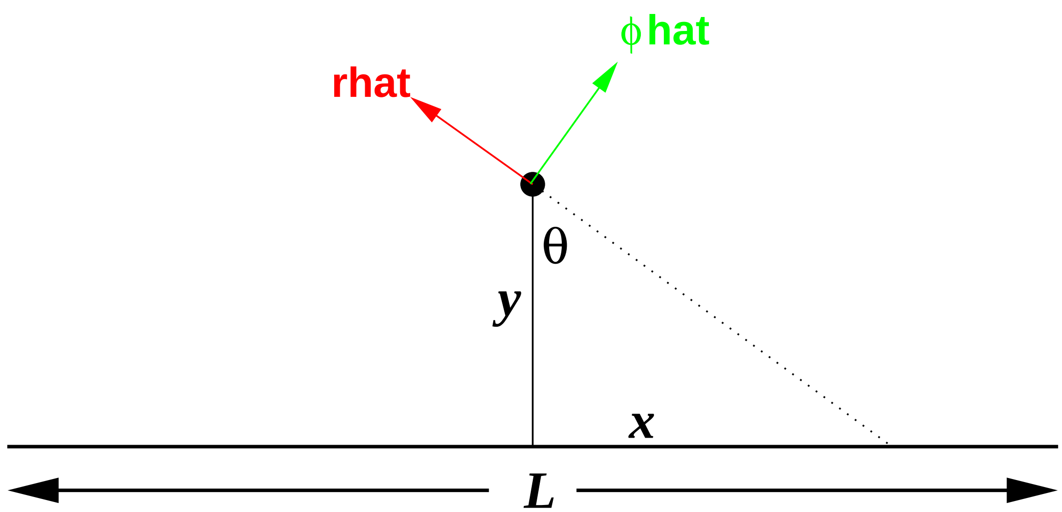
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2004

LECTURE 10: ADDENDUM
CLARIFICATION TO THE MAGNETIC FIELD OF A SHEET

10.7 Field of a plane of current, revisited

The geometry of the vectors that are needed to understand the magnetic field produced by a sheet of current is kind of grotty. A *lot* of people were somewhat befuddled at the end of the March 10 lecture. To clarify things, I present here some more detailed versions of the figure that appeared in Lecture 10; hopefully, this will help lay out why the magnetic field is (a) purely horizontal and (b) depends on the *cosine* of the angle in the figure.

Here is the sheet. In this figure, I've squished all the wires down into a nice, thin sheet. Consider the contribution to the magnetic field arising from a "chunk" of current to the right of the midpoint:

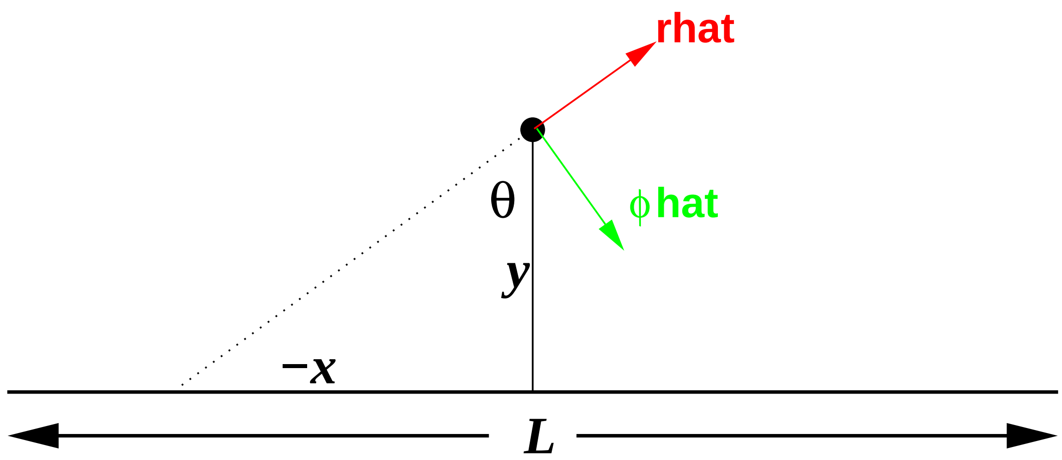


The radial direction from that chunk of current points along the red $\hat{r}$ vector that I've drawn. The magnetic field points along $\hat{\phi}$, which is rotated 90° clockwise (right hand rule) from $\hat{r}$. It is drawn in green. The magnetic field arising from this chunk of current is

$$d\vec{B}_{\rm right} = \frac{2K}{c}\frac{dx}{r}\hat{\phi}_{\rm right}\ ,$$

where $r = \sqrt{x^2 + y^2}$. (Don't forget that the current in this drawing is flowing into the page.)

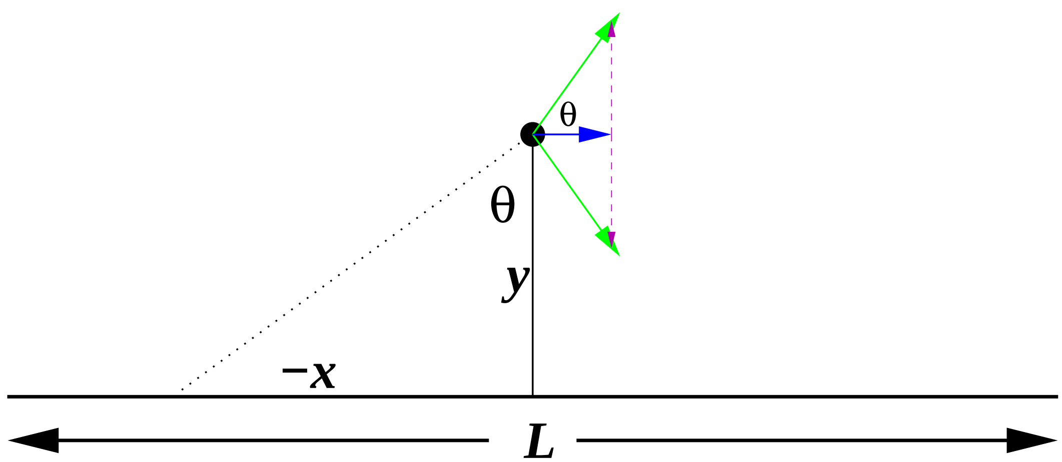
Now consider the field from a chunk to the left of the midpoint:



I've again drawn the radial direction from the chunk of current $\hat{r}$ in red; the magnetic field points along $\hat{\phi}$, rotated 90° clockwise from $\hat{r}$. The magnetic field arising from this chunk of current is

$$d\vec{B}_{\rm left} = \frac{2K}{c}\frac{dx}{r}\hat{\phi}_{\rm left} \;.$$

Now superpose the two contributions: The vertical components (drawn purplish) of the



fields clearly cancel each other out. The horizontal components by contrast reinforce one another. In other words, we can ignore the vertical component of the field and just take the horizontal one.

How do we get the horizontal component? It will clearly be something like $dB_x = |d\vec{B}|$[some trig function](θ); the question is which trig function to use. To get the correct trig function, we need to study the triangles in the figure very carefully — the "little" triangle that is built out of the components of $d\vec{B}$ is congruent with the big one that is built out of x and y. Careful drawing[1] shows that θ belongs where I've drawn it in the figure above.

[1] *Much* more careful than is really possible using thick chalk on a blackboard...

This shows us that the trig function we need is cosine; this in turn tells us that to get the magnetic field we want

$$dB_x = \frac{2K}{c}\frac{dx}{r}\cos\theta \ .$$

Substituting for r and $\cos\theta$ and setting up the integral, we have

$$\begin{aligned}\vec{B} &= \frac{2K}{c}\hat{x}\int_{-L/2}^{L/2}\frac{dx}{\sqrt{x^2+y^2}}\frac{y}{\sqrt{x^2+y^2}} \\ &= \frac{2Ky}{c}\hat{x}\int_{-L/2}^{L/2}\frac{dx}{x^2+y^2} \ .\end{aligned}$$

The rest of the derivation follows from here.

Summary: The key confusing bit is that the horizontal component of the field was given by the cosine of θ (with θ defined as in the drawing). Our usual intuition told us that it should be the sine of θ. However, the intuition we've developed so far is based on the idea that the field direction points along the hypotenuse of the triangle (i.e., along $\hat{r}$). The magnetic field actually points at right angles to the hypotenuse. This extra 90° changes the sine which we might have expected into the cosine that we actually get.

Summary of the summary: magnetic fields can be a pain in the butt.

Scott Hughes *15 March 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 11:
SPACETIME: INTRODUCTION TO SPECIAL RELATIVITY

Preamble: The vast majority of material presented in this course consists of subjects that are challenging, but that I am convinced every 8.022 student can — and should! — learn well. *Special relativity is the only exception to this rule.* The material that we will begin covering today (and that we will conclude on Thursday) is, in my experience, by *far* the most confusing subject that we cover all term. Further, it is not *strictly* necessary to learn this material inside and out in order to understand and excel in 8.022.

You may wonder, given this, why we are bothering to study it at all. The reason is that special relativity "explains" the magnetic field: what we find at the end of all of this is that magnetic forces are nothing but electric forces viewed in another frame of reference. Magnetic fields are just electric fields "in drag", so to speak. Conceptually, this is an *extremely* important point: even if you do not fully grasp the details of where many of the formulas we are about to derive come from, I firmly believe every 8.022 student should understand what magnetic forces "look like" in different frames of reference.

The punchline of this preamble is that you should not be too worried if 8.022's discussion of special relativity makes your brain hurt. Those of you who take more physics courses *will* encounter these concepts several more times; if you don't quite get it now, you will have plenty of chances to get it later[1]. Try to get the flavor of what is happening now, and practice manipulating these formulas, even if it is not 100% clear what is going on.

11.1 Principles of special relativity

Special relativity depends upon two postulates:

1. The laws of physics are the same in all frames of reference.

2. The speed of light is the same in all frames of reference.

These postulates beg the question — what is a frame of reference? A frame of reference (more properly an *inertial* frame of reference) is any set of coordinates that can be used to describe a non-accelerating observer. In the following discussion we will compare measurements in two frames of reference: a station, and a train that is moving through this station in the x direction with constant speed v.

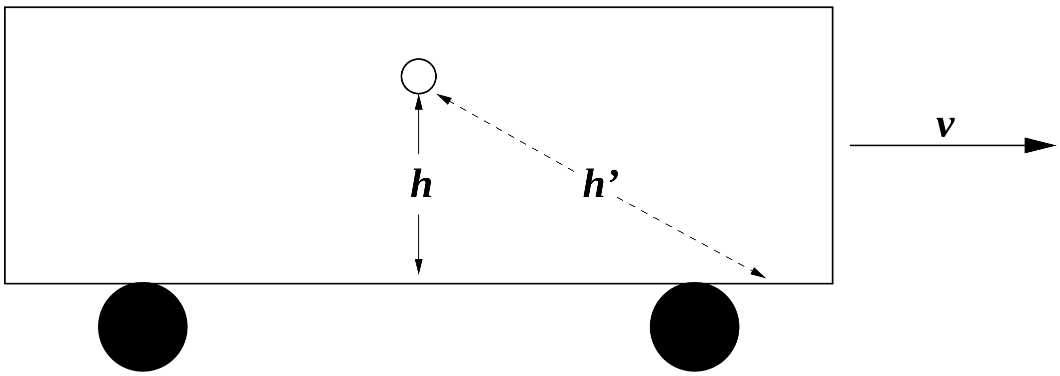
11.2 Time dilation

These postulates have consequences that are rather amazing. In particular, it means that inertial observers in different frames of reference measure *different intervals of time* between

[1] I studied special relativity in 4 separate courses as an undergrad; I honestly can't say I really knew what I was doing until about halfway through the third time. Even then, the fourth study was useful to me.

events, and *different spatial separations* between events. As we'll see in a moment, clocks that appear to be moving run slow; rulers that appear to be moving are shrunk.

Consider a pulse of light emitted from a bulb in the center of a train. The train is moving at speed v in the x direction with respect to a station. The pulse of light is directed toward a photosensor on the floor of the train, directly under the bulb. Two *events* are of interest to observers on the train and in the station: the emission of light from the bulb, and the reception of this light by the photodetector.



Our first question is: How much time passes between the emission of this pulse and its reception on the floor?

The amount of time that is measured to pass is different for observers on the train (at rest with respect to the bulb and the photosensor) than it is for observers in the station (who see the bulb and the photosensor in motion). Let's look at things in the train frame first: the amount of time between emission and reception is just the vertical distance h the light pulse travels divided by the speed of light c:

$$\Delta t_{\text{train}} = \frac{h}{c}$$

How do things look to observers in the station, watching the train go by? They agree that the pulse of light moves vertically through a distance h. However, they claim it also moves through *horizontal* distance because the train is moving with speed v. If the elapsed time between emission and reception as seen in the station frame is $\Delta t_{\text{station}}$, then the pulse moves through a total distance $h' = \sqrt{h^2 + (v\Delta t_{\text{station}})^2}$. Since the speed of light is the same in all frames of reference, the time interval $\Delta t_{\text{station}} = h'/c$:

$$\Delta t_{\text{station}} = \frac{1}{c}\sqrt{h^2 + (v\Delta t_{\text{station}})^2}$$

Substitute $h = c\Delta t_{\text{train}}$:

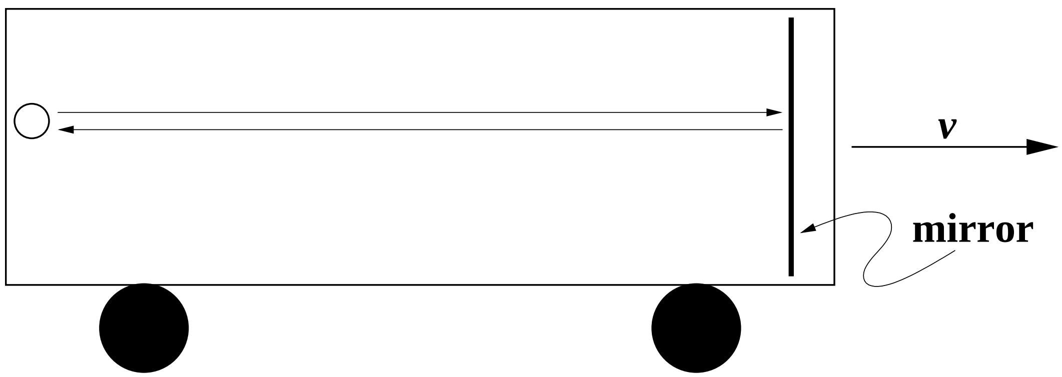
$$(\Delta t_{\text{station}})^2 = (\Delta t_{\text{train}})^2 + \frac{v^2}{c^2}(\Delta t_{\text{station}})^2$$

$$\begin{aligned}\Delta t_{\text{station}} &= \frac{\Delta t_{\text{train}}}{\sqrt{1 - v^2/c^2}} \\ &\equiv \gamma \Delta t_{\text{train}}\end{aligned}$$

where we have defined $\gamma = 1/\sqrt{1 - v^2/c^2}$; note that $\gamma \geq 1$. This relation tells us that *the moving clock runs slow*: the time interval measured by observers who see the clock moving (those in the station) is larger than the time interval measured by observers who are at rest with respect to it.

11.3 Length contraction

Let's move our light source to the back of the train. We flash it in the forward direction, where it reflects off a mirror at the front of the train. The light then returns to the light source, where we measure it. This is a way to measure the length of the railway car.



The two events of interest in this case are the emission of light by the bulb, and its reception after it bounces back to us.

Let's examine what's going on in the train's reference frame first. If the elapsed time between emission of the light and reception is Δt_{train}, we would infer that the length of the train is

$$\Delta x_{\text{train}} = c\Delta t_{\text{train}}/2$$

How do things look in the station frame? We break down the light travel into two pieces: emission to the mirror, then mirror back to reception. Let's say it takes a time $\Delta t_{\text{station, 1}}$ for the first piece — emission to the mirror. An interval $\Delta t_{\text{station, 2}}$ passes following reflection to travel back to reception. If the length of the train as measured in the station is $\Delta x_{\text{station}}$, then the travel times in each bounce is given by

$$\begin{aligned}
\Delta t_{\text{station,1}} &= (\Delta x_{\text{station}} + v\Delta t_{\text{station,1}})/c \\
\Delta t_{\text{station,2}} &= (\Delta x_{\text{station}} - v\Delta t_{\text{station,2}})/c
\end{aligned}$$

The time from emission to mirror is longer since the mirror is moving in the same direction as the light — the light has to chase the mirror on the first piece of the round trip, which takes extra time. On the second piece, the back of the train is rushing towards the light, so it takes less time.

We rearrange these expressions to isolate the travel times:

$$\begin{aligned}
\Delta t_{\text{station,1}} &= \frac{\Delta x_{\text{station}}}{c - v} \\
\Delta t_{\text{station,2}} &= \frac{\Delta x_{\text{station}}}{c + v}
\end{aligned}$$

Then, we add them together to get the total travel time:

$$\begin{aligned}\Delta t_{\text{station}} &\equiv \Delta t_{\text{station},1} + \Delta t_{\text{station},2} \\ &= \frac{2\Delta x_{\text{station}}/c}{1 - v^2/c^2} \\ &= 2\gamma^2 \Delta x_{\text{station}}/c\end{aligned}$$

From our discussion of time dilation, we also know that $\Delta t_{\text{station}} = \gamma \Delta t_{\text{train}} = 2\gamma \Delta x_{\text{train}}/c$. Combining this with our result for $\Delta t_{\text{station}}$, we get

$$\gamma \Delta x_{\text{train}} = \gamma^2 \Delta x_{\text{station}}$$

or

$$\Delta x_{\text{station}} = \Delta x_{\text{train}}/\gamma$$

Since $\gamma \geq 1$, this tells us that *moving objects are contracted*: the train's length according to observers that see it moving is less than the length measured in the "rest frame".

11.3.1 No contraction for perpendicular directions

What about the dimensions of the train that are perpendicular to the train's motion? Without getting too technical, a simple thought experiment can convince us that the perpendicular directions cannot be affected by the motion.

Imagine a train moving towards the mouth of a tunnel. Suppose that the tunnel is 4 meters tall in its rest frame, and that the train is 3.5 meters tall in its rest fram. Suppose the train is moving at 90% of the speed of light. What would happen if the Lorentz contraction shrunk the vertical size of the train?

In the rest frame of the tunnel, the train would appear very short: its height would be Lorentz contracted to $h_{\text{train, contracted}} = 3.5\,\text{meters}/\gamma = 1.53\,\text{meters}$. It fits into the tunnel with no problem.

However, it is equally valid to say that the train is at rest and the tunnel is moving toward it at $0.9c$. Observers in the rest frame of the tunnel claim that the tunnel is shrunk to a height $h_{\text{tunnel, contracted}} = 4.0\,\text{meters}/\gamma = 1.74\,\text{meters}$ — far smaller than the train itself!!!

In one frame, the train fits; in the other, it gets smashed. This is *not* consistent physics! Lorentz contraction of the perpendicular directions does not happen since it does not follow the rules that physics be the same in all reference frames.

11.3.2 Example: Cosmic ray muons

Muons are elementary particles that act, for all practical purposes, like heavy electrons. They are commonly created on earth when high energy protons (cosmic rays) slam into the upper reaches of the earth's atmosphere, several tens of kilometers about the surface. Muons decay into an electron and a neutrino in about 2×10^{-6} seconds. This suggests a mystery: how can a particle that is produced 20 kilometers about the earth possibly reach the ground if it decays in 2 microseconds? Even if the particle were moving at nearly the speed of light, it could only move about 600 meters before decaying.

Special relativity saves the day. Suppose the muon is moving at 99.99% of the speed of light. As seen on the surface of the earth, time dilation stretches out its lifetime to

$$\Delta t_{\text{muon}} = \frac{2 \times 10^{-6}\,\text{sec}}{\sqrt{1-0.9999^2}} = 1.4 \times 10^{-4}\,\text{sec}$$

With this lifetime, the muon can travel a distance

$$\Delta x_{\text{muon}} = 0.9999 \times c \times 1.4 \times 10^{-4}\,\text{sec} = 42\,\text{kilometers}$$

before decaying. The muon can easily make it from high up in the atmosphere to the ground.

It's also interesting to think about this from the muon's perspective. In this frame, the muon is at rest, and the earth's atmosphere is rushing past it at 99.99% of the speed of light. This means that the atmosphere is length contracted by a factor of $1/\gamma = 1/71$. From the muon's perspective, the atmosphere is so thin that it can easily reach the ground.

The two perspectives shown here encapsulate what the slogan "*The laws of physics are the same in all frames of reference*" really means. On the earth's surface, we say that the lifetime of the muon is time dilated until it lives long enough to reach the earth's surface. In the muon's rest frame, its lifetime is only 2 microseconds, so there's no way it could make it through tens of kilometers of atmosphere. However, in this frame, the atmosphere is length contracted until the muon can easily penetrate it in 2 microseconds. The two observers interpret the *details* of what is going on very differently. But both agree on the most important aspect of the phenomena: the muon reaches the ground.

11.4 Lorentz transformation

Length and time dilation are specific manifestations of a *general* consequence of special relativity: what we consider to be "time" and "space" in one reference frame is mixture of the notions of "time" and "space" as measured in another reference frame. The general relation between time and space coordinates in difference frames of reference is described by the *Lorentz transformation.*

The Lorentz transformation is a linear relation between the coordinates of the "rest frame" (for example, the train station) and the "moving frame" (for example, on the train). Let us denote the coordinates of the rest frame as (t, x, y, z); the coordinates of the moving frame are (t', x', y', z'). The moving frame moves with speed v in the x direction, as seen in the rest frame.

The general form of the transformation that relates the two coordinates is

$$\begin{aligned} x' &= Ax + Bt \\ t' &= Cx + Dt \\ y' &= y \\ z' &= z \end{aligned}$$

The y and z coordinates are the same in both frames since the relative motion is in the x direction. Note that the transformation is linear — there are no terms that go proportional to t^2, for example. This is necessary to ensure that the reference frames are inertial — if there were nonlinear terms, the frames would appear to be accelerating with respect to one another rather than moving at a relative velocity.

Our goal now is to calculate A, B, C, and D. We first require that the two coordinate systems must match at the origin: the coordinate $x' = 0$ must match $x = vt$:

$$\begin{aligned} x' &= Ax + Bt \\ &= Avt + Bt \\ &= 0 \end{aligned}$$

The first line is the general Lorentz tranformation relating x' to x and t. On the second line, we plug in $x = vt$. Setting $x' = 0$ and solving, we find.

$$\rightarrow B = -Av \ .$$

Matching at the origin also requires that $x = 0$ must match $x' = -vt'$:

$$\begin{aligned} x' &= A(x - vt) \qquad \text{(General transform for } x'\text{)} \\ &= -Avt \qquad (x = 0) \\ &= -vt' \qquad (x' = -vt') \\ &= -v(Cx + Dt) \qquad \text{(General transform for } t'\text{)} \\ &= -vDt \qquad (x = 0) \end{aligned}$$

Comparing the second and fifth lines, we have

$$A = D$$

OK, so far we've determined two of the four constants, and so we can rewrite the Lorentz transformation as

$$\begin{aligned} x' &= A(x - vt) \\ t' &= Cx + At \end{aligned}$$

To determine the remaining two constants, we think about the properties of light pulses as described by the two coordinate systems.

First, consider a light pulse that is sent out in the x direction. We shoot this pulse out at $t = 0$. The location of the pulse in the unprimed coordinate system is $x = ct$. This pulse will also move out parallel to the x' coordinate system since the relative motion of the two systems is in this direction.

Since the speed of light is the same in all reference frames, its position in the primed coordinate system is given by $x' = ct'$:

$$\begin{aligned} x' &= ct' \\ A(x - vt) &= c(Cx + At) \\ Act - Avt &= c^2Ct + cAt \\ -Av &= c^2C \\ \rightarrow C &= -Av/c^2 \end{aligned}$$

The Lorentz transformation now becomes

$$\begin{aligned} x' &= A(x - vt) \\ t' &= A(t - vx/c^2) \end{aligned}$$

To finally determine A, we shoot a light pulse out along the y axis. In the unprimed coordinate system, the location of the pulse is given by $y = ct$, $x = 0$. What is its location in the primed system? The motion of the frame itself will cause it to move in the x' direction. The total distance that the pulse moves is ct'. We therefore must have $(x')^2 + (y')^2 = c^2(t')^2$:

$$\begin{aligned}
(x')^2 + (y')^2 &= c^2(t')^2 \\
A^2(x - vt)^2 + y^2 &= c^2 A^2(t - vx/c^2)^2 \\
A^2(-vt)^2 + c^2t^2 &= c^2 A^2 t^2 \qquad (x = 0, y = ct) \\
A^2 &= \frac{1}{1 - v^2/c^2} = \gamma^2
\end{aligned}$$

We thus have finally fixed the Lorentz transformation's form:

$$\begin{aligned}
x' &= \gamma(x - vt) \\
t' &= \gamma(t - vx/c^2)
\end{aligned}$$

This can be written more symmetrically by definining $\beta = v/c$ and multiplying the t transformation by c:

$$\begin{aligned}
x' &= \gamma[x - \beta(ct)] \\
ct' &= \gamma[ct - \beta x]
\end{aligned}$$

The inverse of this transformation is quite simple. Observers at rest on the train see the station moving in the $-x$ direction with speed v. This means that the Lorentz transformation is inverted by just reversing the direction of v — i.e., we replace v with $-v$:

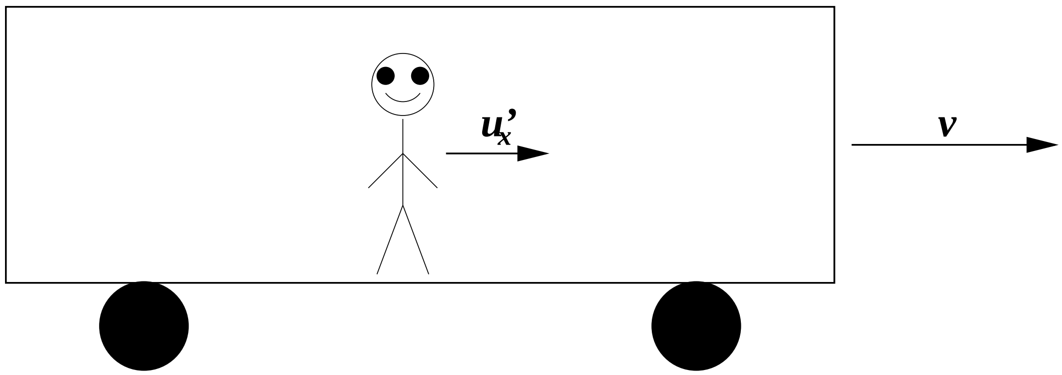
$$\begin{aligned}
x &= \gamma(x' + vt') \\
t &= \gamma(t' + vx'/c^2)
\end{aligned}$$

or

$$\begin{aligned}
x &= \gamma[x' + \beta(ct')] \\
ct &= \gamma[ct' + \beta x']
\end{aligned}$$

11.5 Transformation of velocity

An important result that follows from this is a formula for addition of velocity. Suppose I am in the moving train, walking from one end to the other with a speed $dx'/dt' = u'_x$:



With what speed u_x does someone in the station frame see me move? In old fashioned mechanics, we would just add my speed to the speed of the moving train, and so we'd have $u_x = u'_x + v$. To do it correctly in special relativity, we use the Lorentz transformation:

$$\begin{aligned} u_x &= \frac{dx}{dt} \\ &= \frac{\gamma(dx' + vdt')}{\gamma(dt' + vdx'/c^2)} \\ &= \frac{(dx'/dt') + v}{1 + (v/c^2)(dx'/dt')} \\ &= \frac{u'_x + v}{1 + (vu'_x/c^2)} \end{aligned}$$

This formula shows how it is impossible for anything to move faster than the speed of light. Suppose the train is moving at $v = 0.95c$. Suppose further that my walking speed within the train is $u'_x = 0.95c$. In old fashioned mechanics, we would predict that my walking speed as seen from the ground would be $1.9c$ — almost twice the speed of light. In special relativity, we find

$$\begin{aligned} u_x &= \frac{0.95c + 0.95c}{1 + 0.95 \times 0.95} \\ &= \frac{1.9c}{1.9025} \\ &= 0.9987c \end{aligned}$$

Close to, but still less than, the speed of light.

Suppose that I walk perpendicular to the train's motion: I walk with

$$u'_y = \frac{dy'}{dt'}$$

What is this as seen from the ground? Again, we use the Lorentz transformation:

$$u_y = \frac{dy}{dt}$$

$$
\begin{aligned}
&= \frac{dy'}{\gamma(dt' + v dx'/c^2)} \\
&= \frac{dy'/dt'}{\gamma[1 + (v/c^2)(dx'/dt')]} \\
&= \frac{u'_y}{\gamma[1 + (v u'_x/c^2)]}
\end{aligned}
$$

Likewise,

$$
u_z = \frac{u'_z}{\gamma[1 + (v u'_x/c^2)]}
$$

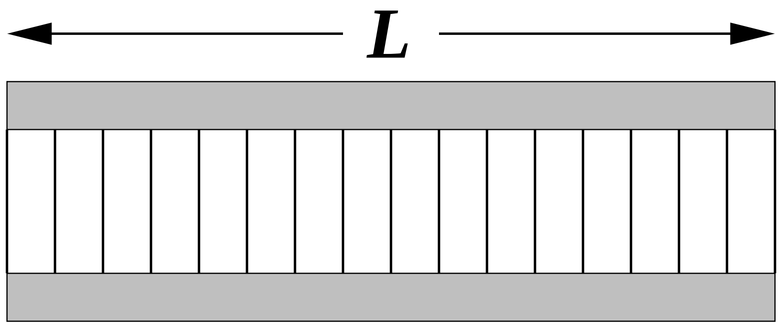
We will use these formulas in the next lecture to explore how things transform from the "viewpoint" of moving charges and currents.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 12:
FORCES AND FIELDS IN SPECIAL RELATIVITY

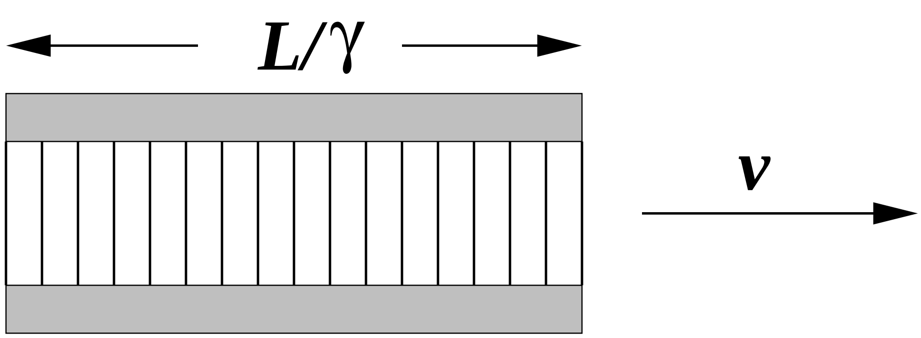
12.1 Transformation of the electric field

It is very easy to see how the electric field transforms by considering a special case. Examine a capacitor in its rest frame. Orient the capacitor such that its plates are parallel to the $x-y$ plane. We take the capacitor to have charge density σ on the lower plate, $-\sigma$ on the upper plate. We take the plates to be square, with each side having length L. The total amount of charge on the plates is thus $Q = \sigma L^2$; the electric field between the plates is $\vec{E} = 4\pi\sigma\,\hat{z}$:



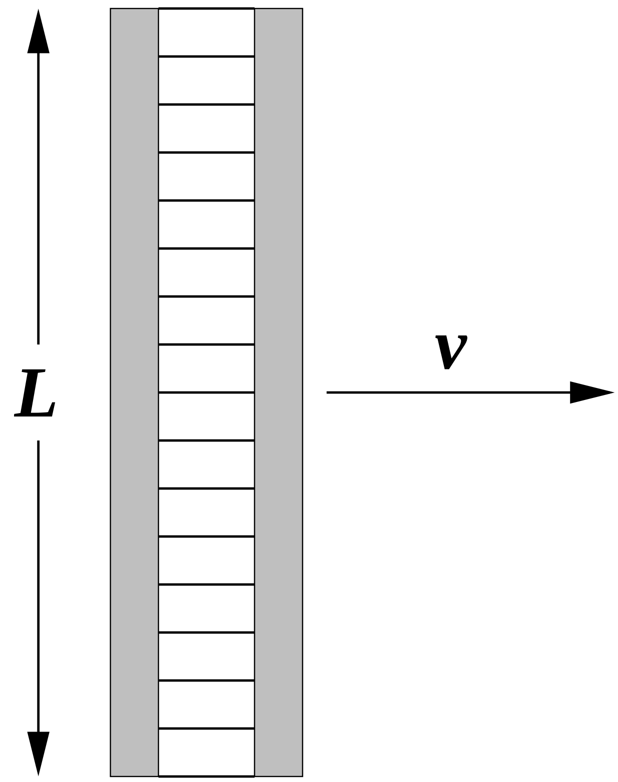
The field inside the capacitor is represented by the field lines sketched here.

Consider now this capacitor "boosted" to a velocity v in the x direction. Charge is a "Lorentz invariant": all observers agree that the total charge on the plates is Q. Because of the Lorentz contraction, the plate's dimension along the direction of motion is seen to be L/γ.



The charge density is thus augmented: $\sigma' = Q/(L \times L/\gamma) = \gamma Q/L^2 = \gamma\sigma$. The electric field is therefore augmented as well: $\vec{E}' = 4\pi\gamma\sigma\,\hat{z}$. Notice that the field lines are denser in this picture: keep in mind that density of field lines tells us how strong the field is.

How about if the capacitor is oriented differently? Let's rotate it so that the plates lie in the $y - z$ plane. The electric field in the rest frame is thus $\vec{E} = 4\pi\sigma\,\hat{x}$. When this capacitor is boosted in the x direction, the plates appear closer together, but the charge density *does not change*: $\sigma' = \sigma$ with this orientation: $\vec{E}' = \vec{E} = 4\pi\sigma$.



Notice that the density of field lines are not changed in this picture, reflecting the fact that the field is not changed by the boost.

Even though we have only looked at the electric fields for a rather special circumstance, the rule we deduce from this is in fact quite general: **Components of the electric field perpendicular to the velocity are augmented by the Lorentz transformation; components that are parallel to the velocity are unaffected**. Mathematically,

$$\begin{aligned} E'_\perp &= \gamma E_\perp \\ E'_{||} &= E_{||}\ . \end{aligned}$$

12.2 Transformation of energy and momentum

Our next goal will be to understand how forces transform between different frames of reference. To do this, we first must understand how to describe energy and momentum in special relativity.

Consider an object whose *rest mass* — the mass that we would measure when this object is at rest with respect to us — is m. If this object moves with velocity $\vec{u}$, what is its energy and its momentum? In the "old-fashioned" mechanics you came to know and love in 8.012, we said that a moving body has a kinetic energy

$$E_{\rm kin} = \frac{1}{2} m |\vec{u}|^2$$

and a momentum

$$\vec{p} = m\vec{u}\ .$$

These definitions form the foundations of kinematics — using them and the notions of *conservation of momentum* and *conservation of energy*, an enormous amount of mechanics follow.

To develop a generalization of kinematics for special relativity, we insist that some notion of "conservation of energy" and "conservation of energy" must be preserved — *some* kind of notion of energy and momentum must be the same before and after two objects collide (for example), even if those objects are moving at near light speed relative to one another. We unfortunately will not be able to go into the "hows" of the way that this works in this course; interested former 8.012 students can find good discussion in Chapter 13 of Kleppner and Kolenkow's textbook; students who take 8.033 next year will study this by the bucketload. I will instead just quote the final result: "conservation of energy" and "conservation of momentum" work in special relativity by requiring that the energy and momentum of a mass observed to move with velocity $\vec{u}$ is given by

$$\begin{aligned} \vec{p} &= \gamma_u m\vec{u} \\ E &= \gamma_u mc^2 \end{aligned}$$

where $\gamma_u = 1/\sqrt{1-u^2/c^2}$. Before moving on, it is worth looking at the small u expansion of these quantities. We expand γ_u as

$$\gamma_u \simeq 1 + \frac{1}{2}\frac{u^2}{c^2} \qquad u \ll c\ .$$

Using this, we find

$$\begin{aligned} \vec{p} &\simeq \left(1 + \frac{1}{2}\frac{u^2}{c^2}\right) m\vec{u} \\ &\simeq m\vec{u} \\ E &\simeq \left(1 + \frac{1}{2}\frac{u^2}{c^2}\right) mc^2 \\ &\simeq mc^2 + \frac{1}{2}mu^2\ . \end{aligned}$$

In the momentum formula, we could have included another term whose magnitude scales proportional to u^3; for small u, this term is negligible and can usually be ignored. The result is nothing more than the "usual" momentum we all learned to love and cherish in 8.012.

The energy formula is particularly interesting: it tells us that the energy of a body with mass m is just the "usual" kinetic energy, $mu^2/2$, plus a "rest energy" mc^2. This is probably the most well-known physics equation on Earth, the energy associated with mass itself.

For our purposes, the most important result will be how these generalizations of energy and momentum Lorentz transform. Suppose in frame 1 — the "unprimed" frame — a body is seen to move with velocity $\vec{u}$. Its energy E and momentum $\vec{p}$ will then be given by the formulas above.

Consider now frame 2. Frame 2 claims that frame 1 is moving with velocity $\vec{v} = v\hat{x}$. What are the energy and angular momentum of the body as seen in frame 2 — the "primed" frame? The Lorentz transformation we need is given by

$$\begin{aligned}
E' &= \gamma_v(E - \beta_v c p_x) \\
p'_x &= \gamma_v(p_x - \beta_v E/c) \\
p'_y &= p_y \\
p'_z &= p_z
\end{aligned}$$

where $\beta_v = v/c$, and $\gamma_v = 1/\sqrt{1 - v^2/c^2}$.

12.3 Transformation of forces

We are now ready to work out how force transforms between reference frames. Let us first work out a few quantities in the "rest frame" of our body. Note that the name "rest frame" is somewhat misleading here — if there's a force acting on it, it must be accelerating, so it isn't going to be at rest for very long! Our "rest frame" will be the frame in which the body is *initially* at rest — $p = 0$ at $t = 0$, and then moves slowly enough that the old-fashioned, slow u description of the body's motion is accurate.

Suppose the force acts in the x direction. In this "rest frame", the force on the body is

$$F_x = \frac{dp_x}{dt} .$$

For what follows, we will also need to know about *changes* in the body's position and in its energy. The change in its position is given by

$$\Delta x = \frac{1}{2}\left(\frac{F_x}{m}\right)\Delta t^2$$

since F_x/m is the acceleration of the body. The change in its energy is given by

$$\Delta E = \frac{(F_x \Delta t)^2}{2m} .$$

You should be able to show very easily that this is just $\frac{1}{2}mu^2$ [bear in mind that $u = a\Delta t = (F_x/m)\Delta t$].

Now consider the way things look in the "lab frame". People in the lab frame see the "rest frame" moving with velocity $v\hat{x}$. They will define the force F'_x as

$$F'_x = \frac{dp'_x}{dt'} .$$

To work this quantity out in terms of F_x we use the Lorentz transformation: we need $\Delta p'_x$ and $\Delta t'$,

$$\begin{aligned}
\Delta p'_x &= \gamma_v\left(\Delta p_x - \beta_v \Delta E/c\right) \\
&= \gamma_v\left[\Delta p_x - \beta_v (F_x\Delta t)^2/(2mc)\right] \\
\Delta t' &= \gamma_v\left(\Delta t - \beta_v \Delta x/c\right) \\
&= \gamma_v\left[\Delta t - \beta_v F_x \Delta t^2/(2mc)\right] \\
&= \gamma_v \Delta t\left[1 - \beta_v F_x \Delta t/(2mc)\right]
\end{aligned}$$

Now, we divide and take a limit:

$$\begin{aligned}
F'_x &= \lim_{\Delta t' \to 0} \frac{\Delta p'_x}{\Delta t'} \\
&= \lim_{\Delta t \to 0} \frac{\gamma_v \left[\Delta p_x - \beta_v (F_x \Delta t)^2/(2mc)\right]}{\gamma_v \Delta t \left[1 - \beta_v F_x \Delta t/(2mc)\right]} \\
&= \lim_{\Delta t \to 0} \frac{\gamma_v \Delta p_x}{\gamma_v \Delta t} \\
&= F_x \ .
\end{aligned}$$

Components of force parallel to the frames' relative motion are the same in both reference frames!

Suppose the force was in a perpendicular direction. The analysis goes through largely the same:

$$\begin{aligned}
F'_y &= \lim_{\Delta t' \to 0} \frac{\Delta p'_y}{\Delta t'} \\
&= \lim_{\Delta t \to 0} \frac{\Delta p_y}{\gamma_v \Delta t \left[1 - \beta_v F_x \Delta t/(2mc)\right]} \\
&= \lim_{\Delta t \to 0} \frac{\Delta p_y}{\gamma_v \Delta t} \\
&= \frac{F_y}{\gamma_v} \ .
\end{aligned}$$

Components of force perpendicular to the frames' relative motion are related by a factor of $1/\gamma$: The force in the "lab" frame is smaller by $1/\gamma$ than the force in the "rest" frame.

12.4 The force exerted by a current upon a moving charge

Warning: the going gets a little rough in this section.

12.4.1 Lab frame

In the lab frame, we have a wire that is electrically neutral, but that carries a current. The wire contains positive charges with density per unit length $\lambda_+ = \lambda_0$. This is the rest frame of those charges, so this means $\lambda_+^{\mathrm{REST}} = \lambda_0$.

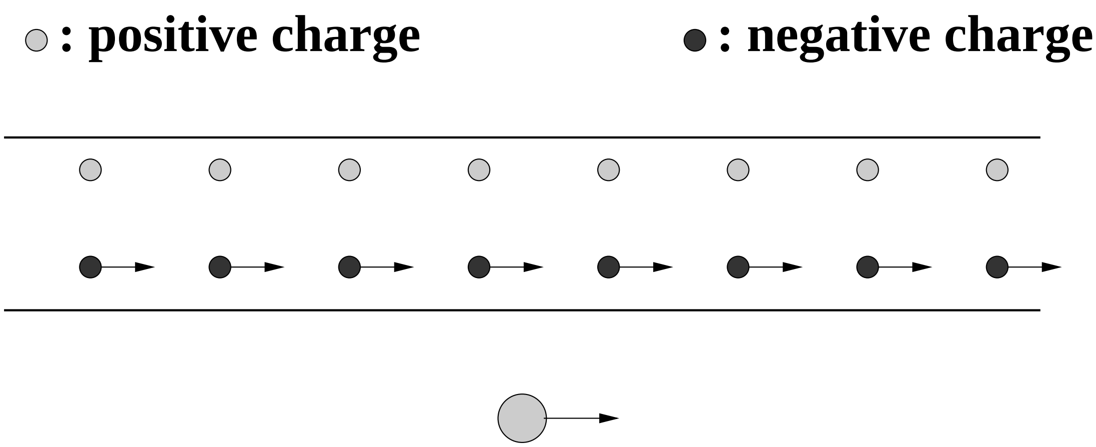


Figure 1: The moving negative charges generate a current moving to the left. This in turn leads to a magnetic field that pushes the charge Q away from the wire.

The wire also contains negative charges; these charges are of course moving with some velocity $\vec{u}$. Since the wire is neutral overall, the density of these negative charges is $\lambda_- = -\lambda_0$. *This is not the density of the electrons in THEIR OWN rest frame!!!!* Suppose that the charge density of the electrons in their own rest frame is $\lambda_-^{\mathrm{REST}}$. By the Lorentz contraction, we must have

$$\lambda_- = \gamma_u \lambda_-^{\mathrm{REST}}$$

where $\gamma_u = 1/\sqrt{1 - u^2/c^2}$. Doing the math, this tells us that

$$\lambda_-^{\mathrm{REST}} = -\frac{\lambda_0}{\gamma_u} \ .$$

We'll need this result shortly.

A charge Q outside of the wire is moving to the right. What forces act on the charge? Since the wire is neutral, there can be no electric force. However, there is a current with magnitude $I = \lambda_0 u$ flowing in the wire. This generates a magnetic field,

$$B = \frac{2I}{rc} = \frac{2\lambda_0 u}{rc}$$

and so there is a force of magnitude

$$F = qvB/c = \frac{2q\lambda_0 uv}{rc^2} \ .$$

By doing the right hand rule game, we quickly see that this force repels the charge from the wire.

12.4.2 Charge's rest frame

By definition, the charge sits still in this frame. Since it's not moving, there *cannot* be a magnetic force that acts on it. But, for there to be a consistent description of the system in the two reference frames there has to be *some kind* of force. What can it be? The only possibility is that there is an *electric* force.

"Where the hell does that come from???" you ask. Let's find out by analyzing the charges in the wire in this frame:

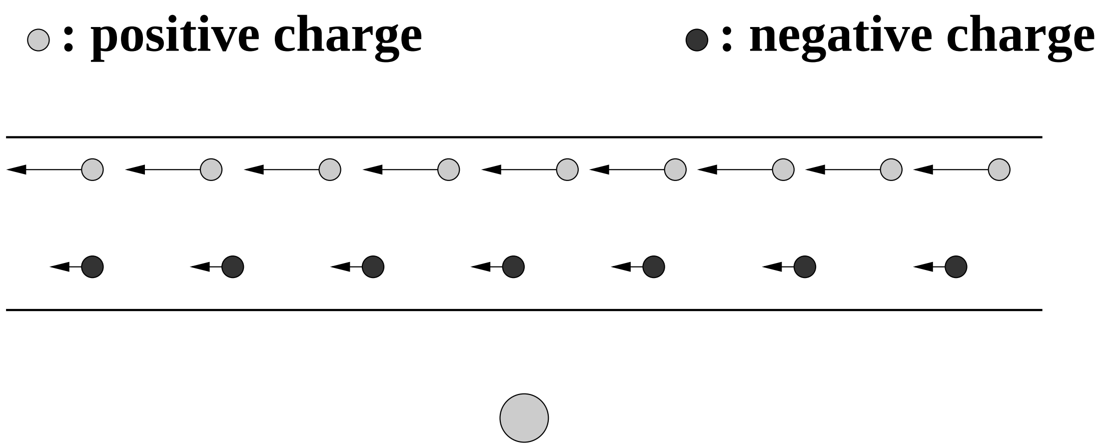


Figure 2: Currents in the wire are irrelevant since the charge does not move in this frame. However, the different length contractions experienced by the positive and the negative charges leads to the wire having a net positive charge as seen in this frame, leading, again, to a force repelling the charge from the wire.

Positive charges: the positive charges are moving in this frame with velocity $-\vec{v}$. Their density as seen in this frame is thus given by

$$\lambda'_+ = \gamma_v \lambda_+^{\mathrm{REST}} = \gamma_v \lambda_0 \ ,$$

where $\gamma_v = 1/\sqrt{1 - v^2/c^2}$.

Negative charges: the negative charges are moving in this frame with a velocity given by adding the velocity $-\vec{u}$ to $\vec{v}$ in the correct, relativistic fashion: their velocity is

$$u' = \frac{u - v}{1 - uv/c^2}$$

and so their density is given by

$$\lambda'_- = \gamma_{u'} \lambda_-^{\mathrm{REST}} = -\frac{\gamma_{u'}}{\gamma_u} \lambda_0 \ .$$

The Lorentz factor $\gamma_{u'}$ is a bit of a mess: we need $1/\sqrt{1 - (u')^2/c^2}$. Let's look at this methodically. First, to simplify the notation a bit, define

$$\begin{aligned} \beta_v &= v/c \\ \beta_u &= u/c \\ \beta_{u'} &= u'/c \ . \end{aligned}$$

Next, look at $1-(u'/c)^2 = 1-\beta_{u'}^2$:

$$\begin{aligned}
1-\beta_{u'}^2 &= 1-\left(\frac{\beta_u-\beta_v}{1-\beta_u\beta_v}\right)^2 \\
&= 1-\frac{\beta_u^2+\beta_v^2-2\beta_u\beta_v}{(1-\beta_u\beta_v)^2} \\
&= \frac{1-2\beta_u\beta_v+\beta_u^2\beta_v^2}{(1-\beta_u\beta_v)^2}-\frac{\beta_u^2+\beta_v^2-2\beta_u\beta_v}{(1-\beta_u\beta_v)^2} \\
&= \frac{1-\beta_u^2-\beta_v^2+\beta_u^2\beta_v^2}{(1-\beta_u\beta_v)^2} \\
&= \frac{(1-\beta_u^2)(1-\beta_v^2)}{(1-\beta_u\beta_v)^2} \\
&= \frac{1}{\gamma_u^2\gamma_v^2(1-\beta_u\beta_v)^2} \ .
\end{aligned}$$

This means that

$$\gamma_{u'} = \gamma_u\gamma_v(1-\beta_u\beta_v) \ .$$

We now finally have enough information to calculate the density of negative charges:

$$\begin{aligned}
\lambda'_- &= -\frac{\gamma_{u'}}{\gamma_u}\lambda_0 \\
&= -\gamma_v(1-\beta_u\beta_v)\lambda_0 \ .
\end{aligned}$$

The *net, total* charge density of the wire in this frame is given by adding the contributions of the negative and the positive charges:

$$\begin{aligned}
\lambda'_{\rm net} &= \lambda'_+ + \lambda'_- \\
&= \gamma_v\lambda_0-\gamma_v(1-\beta_u\beta_v)\lambda_0 \\
&= \gamma_v\beta_u\beta_v\lambda_0 \ . \\
&= \gamma_v\frac{uv\lambda_0}{c^2} \ .
\end{aligned}$$

The wire has a net positive line charge density in this reference frame!!! This means it generates an electric field

$$\begin{aligned}
E' &= \frac{2\lambda'_{\rm net}}{r} \\
&= \frac{2\gamma_v uv\lambda_0}{rc^2}
\end{aligned}$$

which in turn means that there is a force

$$F' = \frac{2\gamma_v q\lambda_0 uv}{rc^2} \ .$$

repelling it from the wire.

Note that there is also a magnetic field in this frame — there are currents associated with both the positive and the negative charges in the wire. However, since the charge Q is at rest, only the electric field is relevant to the force that it experiences.

12.4.3 Comparison of the forces in the two frames

In the "lab frame", we see the charge moving with speed v and deduce that it will feel a magnetic force

$$F = \frac{2uv\lambda_0}{rc^2} \ .$$

In the "charge frame", the charge is at rest, so it cannot feel any magnetic force. However, we just showed that in this frame the wire is not electrically neutral, and so the charge experiences an electric force of

$$F' = \gamma_v \frac{2uv\lambda_0}{rc^2} \ .$$

Are these results consistent? *Yes!!* Remember that rule for transforming forces: the force that is experienced in the "lab frame" is related to that in the moving frame by a factor of $1/\gamma_v$. In other words, we expect to find

$$F = \frac{1}{\gamma_v} F'$$

— *exactly* what we do in fact find.

You may object that one force is electric and the other is magnetic — aren't we comparing apples and oranges here? For many years, people thought this way: electric forces were one phenomenon, magnetic forces were another. There was no connection between the two. As thought experiments like this show, however, this distinction is a false one. The electric force acting on a charge in that charge's rest frame is *exactly* what we need to explain the magnetic force in a frame in which that charge is moving.

Physics is consistent: even though we give different detailed explanations ascribing what mechanism produces the forces in the two reference frames, we agree *exactly* as to what this force should be. This is the essence of special relativity! It also tells us that electric forces and magnetic forces are really the same thing. "Electricity" and "magnetism" are not separate phenomena: they are different specific manifestations of a single critter, "electromagnetism".

12.5 Summary

The discussion of the past two lectures is probably the most difficult that we will face all semester. Let's recap the major points:

- **Moving clocks run slow.** Time dilation means that a clock which is moving very quickly relative to you will tick more slowly than a clock that is at rest with respect to you.

- **Moving rulers are shortened.** The Lorentz contraction means that you will measure a car shooting past you at very high speed to be shorter than an identical car that is at rest with respect to you.

  - **But only along the direction of motion!!!** There is no Lorentz contraction in a direction orthogonal to the relative motion.

- **What looks like a pure magnetic field in one frame looks like a mixture of magnetic and electric fields in another frame.** This guarantees that there is some kind of force acting on our charge even when we go into its rest frame.

Scott Hughes *29 March 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 13:
AMPERE'S LAW REVISITED; BIOT-SAVART LAW
VECTOR POTENTIAL

13.1 Return to the magnetic field

13.1.1 What we did before

Let's quickly recap the major facts we've gone over regarding the magnetic field. A charge q moving with velocity $\vec{v}$ in a magnetic field $\vec{B}$ feels a force

$$\vec{F} = q\frac{\vec{v}}{c} \times \vec{B} \; .$$

Magnetic fields arise from flowing currents. The integral of $\vec{B}$ along a closed path C the is related to the current enclosed by the path by Ampere's law:

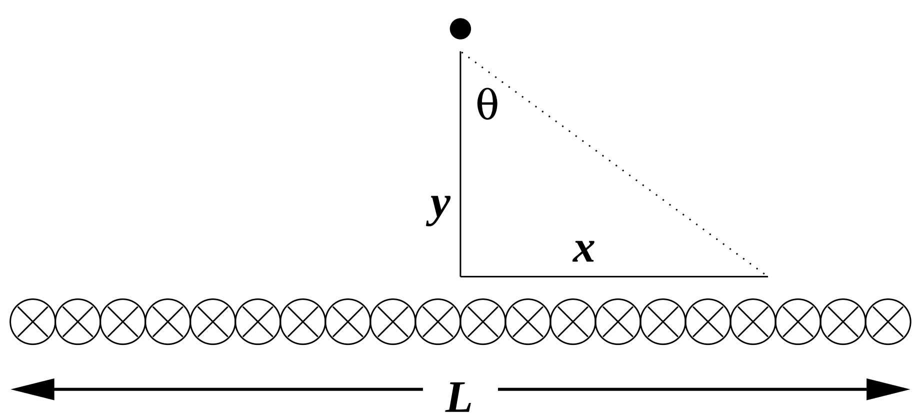
$$\oint_C \vec{B} \cdot d\vec{s} = \frac{4\pi}{c} I_{\rm encl} \; .$$

If we apply this law to a long wire we find

$$\vec{B}(r) = \frac{2I}{cr}\hat{\phi} \; .$$

where r is the distance to the wire and $\hat{\phi}$ is related to the direction in which the current flows by the right-hand rule: point your right-hand thumb in the direction in which the current flows and $\hat{\phi}$ curls around in the direction of your fingers.

By placing many long wires next to each other and invoking superposition, we found the magnetic field of a sheet of current:



In the limit of $L \to \infty$, but holding the current per unit length in the sheet $K \equiv I/L$ constant, we found that

$$\vec{B} = +\frac{2\pi K}{c}\hat{x}$$

above the plane, and

$$\vec{B} = -\frac{2\pi K}{c}\hat{x}$$

below the plane. The key observation here is that the *change* in the magnitude of the magnetic field as we cross the current is given by

$$|\Delta\vec{B}| = \frac{4\pi K}{c} \ .$$

This should be reminiscent of the change in electric field when we cross a sheet of charge, $|\Delta\vec{E}| = 4\pi\sigma$.

Taking the divergence of the wire's $\vec{B}$-field, we find

$$\vec{\nabla}\cdot\vec{B} = 0 \ .$$

Although we have only shown this holds for a specific example, this rule holds for all magnetic fields that arise from currents. Physically, it tells us that — as far we can tell — there is no such thing as an isolated magnetic charge. The smallest "chunks" of magnetic field that we can find come in a dipole configuration, with "north" and "south" poles. We never find an isolated "north" pole, with radial field lines.

13.1.2 Ampere's law revisited

Using Stoke's theorem, we can rewrite the line integral of the $\vec{B}$ field in terms of its curl:

$$\oint_C \vec{B}\cdot d\vec{s} = \int_S (\vec{\nabla}\times\vec{B})\cdot d\vec{a} \ .$$

S is any surface bounded by the curve C. This is a nice result, since the current enclosed by C is given by an integral through the *same* surface S:

$$I_{\rm encl} = \int_S \vec{J}\cdot d\vec{a} \ .$$

Putting these two results together, Ampere's law becomes

$$\begin{aligned}\int_S (\vec{\nabla}\times\vec{B})\cdot d\vec{a} &= \frac{4\pi}{c}\int_S \vec{J}\cdot d\vec{a} \\ \int_S \left(\vec{\nabla}\times\vec{B} - \frac{4\pi}{c}\vec{J}\right)\cdot d\vec{a} &= 0 \ .\end{aligned}$$

We finally obtain

$$\longrightarrow \vec{\nabla}\times\vec{B} = \frac{4\pi}{c}\vec{J} \ .$$

This last equation is also called Ampere's law — it expresses the same physics as $\oint_C \vec{B}\cdot d\vec{s} = 4\pi I/c$, but in differential rather than integral form. If we had been doing this analysis in SI units rather than cgs, we would have ended up with

$$\longrightarrow \vec{\nabla}\times\vec{B} = \mu_0\vec{J} \ .$$

It's worthwhile to pause at this moment and collect the various divergence and curl equations for $\vec{E}$ and $\vec{B}$ that we have so far:

$$\begin{aligned}\vec{\nabla}\cdot\vec{E} &= 4\pi\rho\,, \qquad \vec{\nabla}\cdot\vec{B}=0\\ \vec{\nabla}\times\vec{E} &= 0\,, \qquad \vec{\nabla}\times\vec{B}=\frac{4\pi}{c}\vec{J}\,.\end{aligned}$$

These are Maxwell's equations for static fields and steady currents. They are what you see on the t-shirts whenever all time derivatives go to zero. Their SI versions are

$$\begin{aligned}\vec{\nabla}\cdot\vec{E} &= \rho/\epsilon_0\,, \qquad \vec{\nabla}\cdot\vec{B}=0\\ \vec{\nabla}\times\vec{E} &= 0\,, \qquad \vec{\nabla}\times\vec{B}=\mu_0\vec{J}\,.\end{aligned}$$

13.2 The vector potential

13.2.1 Definition

Recall that for static electric fields, we have

$$\vec{E} = -\vec{\nabla}\phi\,.$$

We motivated this rule from arguments based on work and energy — namely, that the potential difference between any two points is just the work it takes (per unit charge) to move a charge between those two points. *However*, it has an important and useful side effect: it "guarantees" that our electric fields are curl free:

$$\vec{\nabla}\times\vec{E} = 0$$

since

$$\vec{\nabla}\times\vec{\nabla}f = 0$$

for any scalar function f.

Can we do anything similar with the magnetic field? In this case, the equation which we would like to enforce is the rule

$$\vec{\nabla}\cdot\vec{B} = 0\,.$$

Can we pick a form of $\vec{B}$ that guarantees this equation holds? Well, obviously, or I wouldn't have wasted the time to type this up: as you proved on Pset 3, the divergence of the curl of any function is zero. This suggests that we can write $\vec{B}$ as the curl of some vector function $\vec{A}$:

$$\vec{\nabla}\times\vec{A} = \vec{B}\,.$$

The function $\vec{A}$ is called the "vector potential". The key thing to bear in mind is that it is introduced to enforce the rule $\vec{\nabla}\cdot\vec{B}=0$, in the same way that $\vec{E}=-\vec{\nabla}\phi$ enforces $\vec{\nabla}\times\vec{E}=0$. Unlike the scalar potential ϕ, however, $\vec{A}$ has *no connection* to work or energy. For our purposes at least, its only use is to facilitate some of our calculations[1].

[1]It turns out to have a very real physical meaning in quantum mechanics, though.

13.2.2 Non-uniqueness

In electrostatics, we were able to prove a uniqueness theorem: once boundary conditions were set, there was one — and *only* one — potential for a given charge distribution, meaning that for every $\vec{E}$ field there was only one potential ϕ.

This is *not* the case for the vector potential! It is actually the case that an *infinite* number of vector potential functions correspond to a particular magnetic field $\vec{B}$. Consider a uniform magnetic field: $\vec{B} = B_0\hat{z}$. Our only requirement on the vector potential $\vec{A}$ is that its curl reproduce this field. In other words, we need

$$\begin{aligned} B_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0 \\ B_y &= \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = 0 \\ B_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B_0 \; . \end{aligned}$$

One vector potential which satisfies these conditions is

$$\vec{A} = xB_0\hat{y} \; .$$

Another is

$$\vec{A} = -yB_0\hat{x} \; .$$

Still another is given by combining these:

$$\vec{A} = -\frac{1}{2}yB_0\hat{x} + \frac{1}{2}xB_0\hat{y} \; .$$

These are *all* valid choices. Given this freedom, it might seem like there's no point in selecting any particular solution. In fact, we end up taking the opposite viewpoint: Since any one of these solutions is valid, let's pick the one that is — in some manner that we must specify — particularly convenient. In the next subsection, we describe a particular choice that allows us to write down an equation that looks a lot like Poisson's equation, but that holds for $\vec{A}$.

13.2.3 Poisson's equation for $\vec{A}$

When we plugged the definition $\vec{E} = -\vec{\nabla}\phi$ into Gauss's law, $\vec{\nabla} \cdot \vec{E} = 4\pi\rho$, we ended up with Poisson's equation,

$$\nabla^2\phi = -4\pi\rho \; .$$

What happens if we plug $\vec{\nabla} \times \vec{A} = \vec{B}$ into Ampere's law, $\vec{\nabla} \times \vec{B} = 4\pi\vec{J}/c$? First, we need to work out the curl of the curl: with a fair amount of tedious algebra you should be able to to show that

$$\vec{\nabla} \times \vec{\nabla} \times \vec{A} = \vec{\nabla}\left(\vec{\nabla} \cdot \vec{A}\right) - \nabla^2\vec{A} \; .$$

This is in principle a horrible mess. However, we can take advantage of the non-uniqueness of $\vec{A}$ to simplify it. We have the freedom to *choose* our $\vec{A}$ in such a way that it has no divergence:

$$\vec{\nabla} \cdot \vec{A} = 0 \; .$$

This is called a "gauge choice". Some of you will study this notion in *far* greater detail in future courses. For now, you simply need to take away the notion that the non-uniqueness of $\vec{A}$ gives us enough freedom to tinker around until the condition $\vec{\nabla} \cdot \vec{A} = 0$ is guaranteed to be met. (It still leaves us a lot of freedom. You can easily verify that all three of the choices for $\vec{A}$ we wrote down in the previous subsection satisfy $\vec{\nabla} \cdot \vec{A} = 0$.)

Imposing this condition, we then have

$$\vec{\nabla} \times \vec{\nabla} \times \vec{A} = -\nabla^2 \vec{A} \,.$$

Let's now plug this into Ampere's law: we have

$$\begin{aligned} \vec{\nabla} \times \vec{B} &= \frac{4\pi}{c} \vec{J} \\ \vec{\nabla} \times \vec{\nabla} \times \vec{A} &= \frac{4\pi}{c} \vec{J} \\ \longrightarrow \nabla^2 \vec{A} &= -\frac{4\pi}{c} \vec{J} \,. \end{aligned}$$

This is kind of a goofy equation. In cartesian coordinates, it is best understood component by component. Basically, it means that each component of $\vec{A}$ is related to a component of $\vec{J}$ by the Laplacian operator:

$$\begin{aligned} \nabla^2 A_x &= -\frac{4\pi}{c} J_x \\ \nabla^2 A_y &= -\frac{4\pi}{c} J_y \\ \nabla^2 A_z &= -\frac{4\pi}{c} J_z \,. \end{aligned}$$

Let's look back at electrostatics again for a moment. Poisson's equation, $\nabla^2 \phi = -4\pi\rho$, has as its solution the general equation for potential,

$$\phi = \int_V \frac{\rho\, dV}{r} \,.$$

Our Poisson equation for the vector potential is the *exact same thing*, provided we replace ϕ with $\vec{A}$ and ρ with $\vec{J}/c$:

$$\vec{A} = \frac{1}{c} \int_V \frac{\vec{J}\, dV}{r} \,.$$

For a current I flowing in a wire, this equation becomes

$$\vec{A} = \frac{I}{c} \int_{\text{along wire}} \frac{d\vec{l}}{r} \,,$$

where $d\vec{l}$ points along the wire.

13.3 The Biot-Savart law

Using the last result we derived — the vector potential of wire — we can work out another very important result. By taking the curl of that vector potential, we derive the magnetic field of a wire:

$$
\begin{aligned}
\vec{B} &= \vec{\nabla} \times \frac{I}{c} \int_{\text{along wire}} \frac{d\vec{l}}{r} \\
&= \frac{I}{c} \int_{\text{along wire}} \vec{\nabla} \times \frac{d\vec{l}}{r} .
\end{aligned}
$$

We can take the "$\vec{\nabla}\times$" under the integral since the limits of integration do not change when we take the derivative. Next, we use

$$
\vec{\nabla} \times \frac{d\vec{l}}{r} = \left(\frac{1}{r}\right) \vec{\nabla} \times d\vec{l} + \vec{\nabla}\left(\frac{1}{r}\right) \times d\vec{l} .
$$

(This follows from stuff you did on pset 3.) $\vec{\nabla} \times d\vec{l} = 0$ since $d\vec{l}$ is a just a differential length along the wire, and $\vec{\nabla}(1/r) = -\hat{r}/r^2$. We find

$$
\vec{\nabla} \times \frac{d\vec{l}}{r} = d\vec{l} \times \frac{\hat{r}}{r^2}
$$

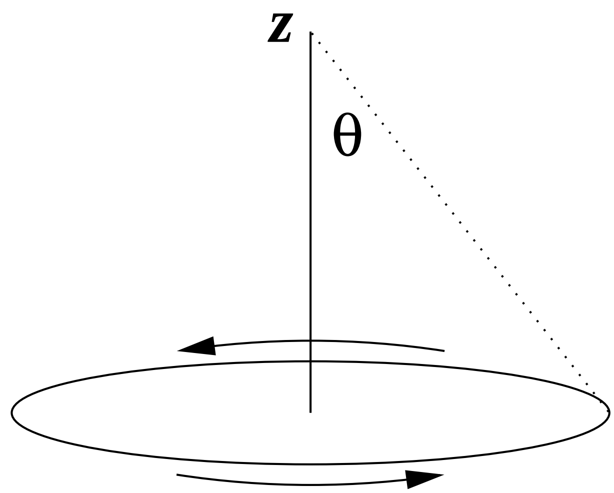
where we've used $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$ to get rid of the minus sign. We thus finally can write down the magnetic field of the wire:

$$
\vec{B} = \frac{I}{c} \int_{\text{along wire}} \frac{d\vec{l} \times \hat{r}}{r^2} .
$$

This result is known as the *Biot-Savart* formula for the magnetic field. It plays roughly the same role in determining the magnetic field from a current that Coulomb's law plays setting the electric field from charges.

13.3.1 Field of a ring of current

The Biot-Savart law allows us to work out the magnetic field of an interesting special case: a ring of current.



Take the ring to have radius R, and consider a current of magnitude I circulating as shown. What is the magnitude and direction of the current a point z above the ring's center?

Set up Biot-Savart:

$$\vec{B} = \frac{I}{c}\int_{\rm around\ ring} \frac{d\vec{l}\times\hat{r}}{r^2}\ .$$

The distance r is easy — $r = \sqrt{z^2+R^2}$. Likewise, the magnitude $|d\vec{l}|$ is easy — $|d\vec{l}| = Rd\phi$, where ϕ is an angle around the perimeter of the ring. The cross product $d\vec{l}\times\hat{r}$ is a little bit trickier. A good way to understand it is to consider a special case: when $z = 0$, $d\vec{l}\times\hat{r}$ points in the vertical $\hat{z}$ direction. Moving away from this case, we see that it *always* points along $\hat{z}$, but is reduced by a factor of $\sin\theta$ in general:

$$\begin{aligned}\vec{B} &= \frac{I}{c}\int_0^{2\pi}\frac{R\sin\theta\,d\phi}{R^2+z^2}\hat{z}\\ &= \frac{I}{c}\int_0^{2\pi}\frac{R^2\,d\phi}{(R^2+z^2)^{3/2}}\hat{z}\\ &= \frac{2\pi I}{c}\frac{R^2}{(R^2+z^2)^{3/2}}\hat{z}\ .\end{aligned}$$

Notice that the direction, $\hat{z}$, could have been predicted by the right-hand rule: curl your fingers of your right hand in the direction of the current, and your thumb points along the field. Note also that it points in the same direction above and below the loop.

13.3.2 The solenoid

Suppose we stack a bunch of rings like this on top of one another. What is the magnetic field at the center of the cylinder that we make?

We answer this by adding up the magnetic fields from many rings. Let the total height of the stack be L. Further, let the total amount of current circulating through *all* of the rings be I. The total current per unit length flowing through the stack is thus $K = I/L$. The contribution to the magnetic field from a height z above the center and from a ring of thickness dz is then

$$\begin{aligned} d\vec{B} &= \frac{2\pi}{c}\frac{I\,dz}{L}\frac{R^2}{(R^2+z^2)^{3/2}}\hat{z} \\ &= \frac{2K\pi}{c}\frac{R^2\,dz}{(R^2+z^2)^{3/2}}\hat{z} \end{aligned}$$

The *total* magnetic field is given by integrating this over z:

$$\vec{B} = \frac{2\pi K}{c}\hat{z}\int_{-L/2}^{L/2}\frac{R^2\,dz}{(R^2+z^2)^{3/2}} \; .$$

This integral turns out to be quite easy to do: since

$$\int_{-L/2}^{L/2}\frac{R^2\,dz}{(R^2+z^2)^{3/2}} = \frac{2L}{\sqrt{L^2+4R^2}}$$

the answer is just

$$\vec{B} = \frac{4\pi K}{c}\frac{L}{\sqrt{L^2+4R^2}}\hat{z} \quad \to \frac{4\pi K\hat{z}}{c} \qquad \text{as } L\to\infty \; .$$

This configuration is known as a "solenoid". The last formula (the $L \to \infty$ limit) is accurate enough (provided $L \gg R$) that it is often used even for finite length solenoids.

This value turns out to describe the magnetic field *throughout* the interior for $L \gg R$! Note that the value is exactly what we got for change of field when we cross a current sheet: in a very long solenoid, there is no field on the outside, but it suddenly jumps to $4\pi K/c$ on the inside. Because of this, solenoids play a role with for magnetic fields similar that played by parallel plate capacitors for electric fields: they are devices which "contain" the field in some isolated region.

In practice, solenoids are constructed by wrapping wires around a tube to make a cylindrical geometry. If we can wrap our wires n times per unit length (for example, we may have $n = 10$ turns/cm), then $K = nI$.

Scott Hughes *31 March 2005*

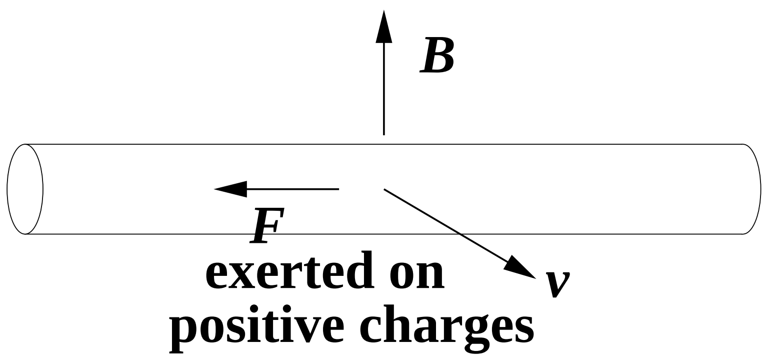
Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 14:
Induction:
Faraday's & Lenz's laws

14.1 Moving wires in magnetic fields

14.1.1 Single length

Suppose we take a length of conducting wire and drag it through a uniform magnetic field:



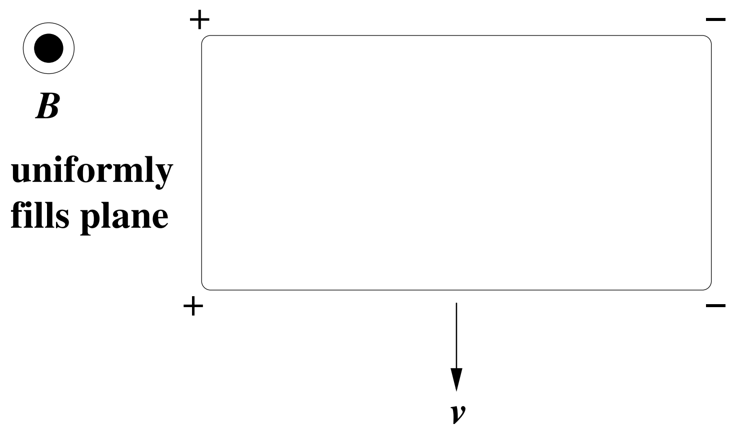
Charges in the wire will feel a force due to being dragged along through the $\vec{B}$-field. This leads to a separation of charges in this rod: as drawn, the left end of the rod will acquire a net + charge, the right end acquires a net − charge.

How much charge separation occurs? By moving the charges apart, we must create an electric field. The electric field will grow until the force that it exerts on the charges balances the magnetic force:

$$q\vec{E}_{\rm balance} + q\frac{\vec{v}}{c} \times \vec{B} = 0$$
$$\longrightarrow E_{\rm balance} = -\frac{\vec{v}}{c} \times \vec{B}\ .$$

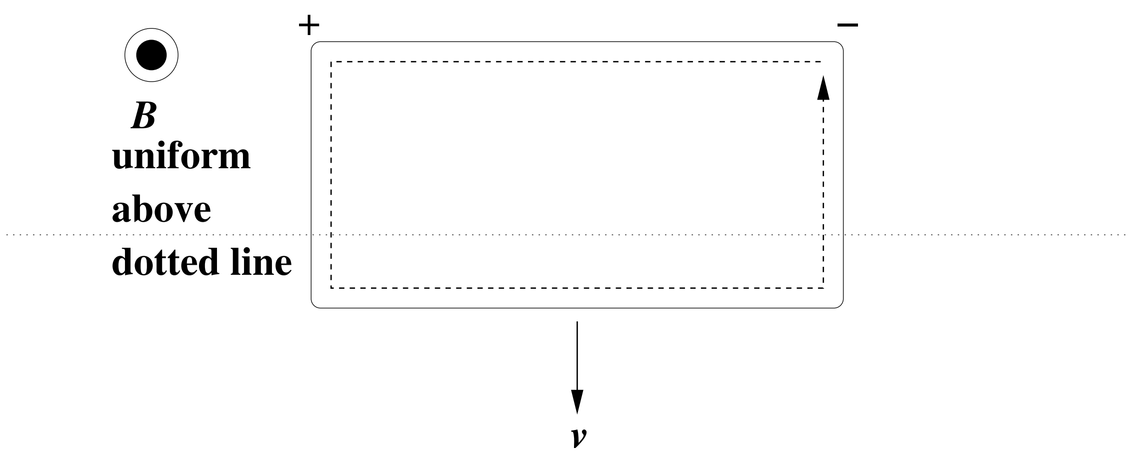
14.1.2 Closed loop

Suppose we now take our wire and bend it into a square. What happens when we drag it through the magnetic field? Well, if the magnetic field is truly uniform and filling all of space, nothing very interesting happens:



We get charge separation in both the top and the bottom legs, so there's an $\vec{E}$-field pointing to the right along both legs. This stops the flow of charge — once the $\vec{E}$-field is built up, the charge separation doesn't do anything else. It's done.

Things get more exciting if we make the magnetic field non-uniform. Suppose that the magnetic field is zero below some line, but is constant and non-zero above that line:



Because of the magnetic force, charge will flow — as we already know — to the left across the upper side of the loop. However, there's now no opposing $\vec{E}$-field that will prevent it from *continuing* to flow! The charges will very happily circumnavigate the entire loop, forming a cyclical current. We say that we have *induced* a current to flow in this loop. This is our first encounter with the phenomenon of electromagnetic induction.

A few other things are worth noticing. First, note that

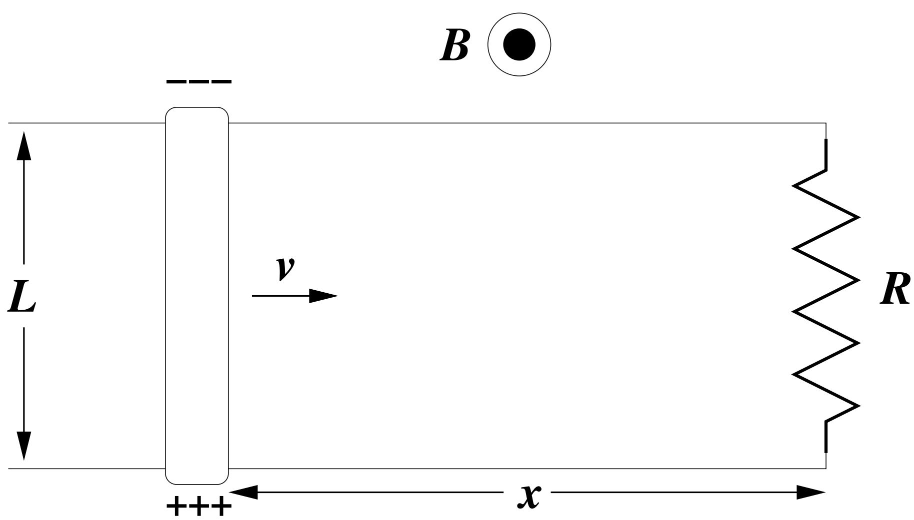
$$\oint_{\text{wire loop}} \vec{E} \cdot d\vec{s} \neq 0 \ .$$

For this to be true, we *cannot* have $\vec{\nabla} \times \vec{E} = 0$! Don't be too disturbed by this — all this tells us is that we have begun to move beyond electrostatics.

Second, notice that the current flowing in this loop will itself feel a force from the magnetic field: the leg across the top of the loop carries a current to the left. Doing the cross product for $\vec{I} \times \vec{B}$, we find that this generates a force on the loop that points up — *opposing* the motion of the loop. (There is also a force to the left on the left-hand side of the loop, and a force to the right on the right-hand side, but these are equal and opposite. The force on the top of the loop is unopposed.) This opposition of the loop's motion is our first contact with a more general rule called *Lenz's law*, which is a kind of electromagnetic inertia.

14.2 Induced EMF

This little example may give you an idea for how to make a simple little generator for electric power. Imagine sliding a bar across a pair of rails, sitting in a region of constant magnetic field:



The motion of the bar causes charge separation, which in turn will drive a current to flow around the loop and through the resistor. This is the same behavior we would have seen if the current were driven by a battery rather than by the moving rod! We can thus think of the motion of the rod as acting like a source of EMF.

What is the value of this EMF? For a battery, we defined the EMF as the potential of the positive terminal with respect to the negative terminal. Bearing in mind that potential is the work (per unit charge) it takes to move charges from point to point, it is simple to compute the EMF of this arrangement:

$$\begin{aligned}
\mathcal{E} &= \frac{1}{q} W(\text{moving from } - \rightarrow +) \\
&= \frac{1}{q} \int_{-}^{+} \vec{F} \cdot d\vec{s} \\
&= \frac{1}{c} \int_{0}^{L} \left(\vec{v} \times \vec{B}\right) \cdot d\vec{s} \\
&= \frac{vBL}{c} \,.
\end{aligned}$$

The current that flows through this loop is then given by

$$I_{\rm ind} = \frac{\mathcal{E}}{R} = \frac{vBL}{cR}\ .$$

14.3 A different way to think about this result

The formula we just derived gives us an EMF that is proportional to $v = dx/dt$, the speed of the bar sliding along the rails. The magnetic flux in the rectangle defined by the rails, the bar, and the resistor is given by

$$\Phi_B = Blx\ .$$

Because of the motion of the rod, this flux decreases with time: taking the derivative, we have

$$\frac{\partial \Phi_B}{\partial t} = -Blv\ .$$

This is (almost) exactly what we just got for the EMF generated by the rod's motion:

$$\mathcal{E} = -\frac{1}{c}\frac{\partial \Phi_B}{\partial t}\ .$$

As we'll discuss in a moment, this formula turns out to be the most general formula we need to describe induced EMF.

For now, the most important thing to understand about this formula is the minus sign. The minus tells us about the direction of the induced EMF; or, if you prefer, the direction in which a current driven by this induced EMF tends to flow. The current flows in such a direction that it *opposes the change in flux* through the loop. This is Lenz's law: "I don't want the flux through me to change, dammit[1]". Lenz's law expresses a kind of inertia for magnetic flux.

Let's examine this carefully. Suppose the bar were sitting still. The flux out of the page is some positive number, since $\vec{B}$ points out of the page, and the area is oriented so that its normal is likewise out of the page (by right-hand rule). Now, we make the bar slide to the right. The *outward flux gets smaller* since the area is decreasing in size. Lenz's law tells us that the current from the induced EMF will fight this change: to oppose the decrease in flux, the current flows in such a way as to augment it. This means it must flow counterclockwise: by right hand rule, a clockwise circulating current will tend to increase the magnetic field pointing out of the page.

If the bar slides to the left, the outward flux would *grow*. Lenz's law wants to fight that too, so it induces a current that will tend to decrease the magnetic field out of the page — a current circulating *clockwise*.

If a current flows in a magnetic field, it feels a force. In this example of a sliding bar, Lenz's law guarantees that this force acts to oppose the bar's motion — it ends up acting like a drag force. Consider the bar sliding to the right: the current circulates counter clockwise, so the $\vec{I} \times \vec{B}$ force points to the left, opposing the bar's motion. Consider the bar sliding to the left: the current circulates clockwise, so the $\vec{I} \times \vec{B}$ force points to the right — again,

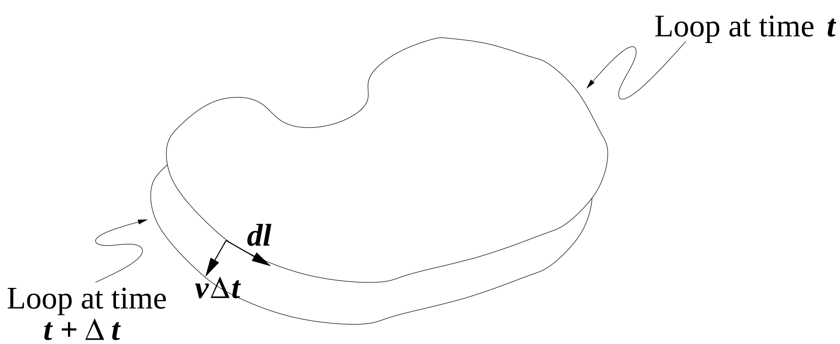
[1]The final "dammit" is in honor of Professor Ray Ashoori, who, during the Spring 2003 semester, was apparently unable to state Lenz's law without a mild expletive.

opposing the bar's motion. In both cases, the interaction of the current with the magnetic field acts as a drag. The minus sign in the EMF equation was *key* to this: if it had been a + sign, we would have found "anti-drag"! Rather than acting to slow down the bar, the magnetic field and magnetic induction would act to *speed it up*!!! You should be able to convince yourself that this is absurd: you quickly end up with a "runaway" configuration, where the bar's speed just increases without limit. This would end up violating conservation of energy. Lenz's law and the minus sign can thus be regarded as enforcing conservation of energy in magnetic induction.

14.4 EMF and $\partial\Phi_B/\partial t$: General proof for static fields

We showed for a specific example that the EMF generated by moving things around in a magnetic field is given by $-(1/c)\partial\Phi_B/\partial t$. We now prove that this result in fact holds in general.

To do so, consider a loop of arbitrary shape. Take this loop to be moving with velocity $\vec{v}$ through an arbitrary — but static — magnetic field $\vec{B}$:



At time t, the magnetic flux through this loop is

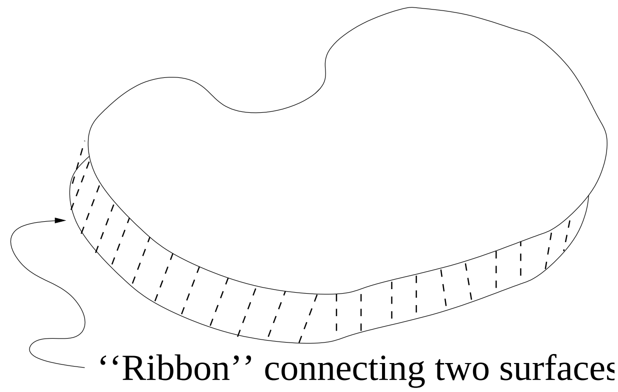
$$\Phi_B(t) = \int_S \vec{B} \cdot d\vec{a}$$

where S is any surface spanned by the wire at time t.

How does the flux change when it moves to its new position at $t + \Delta t$? By looking at the picture, we can see that

$$\Delta\Phi_B = \Phi_B(t + \Delta t) - \Phi_B(t)$$

is just given by the flux through a "ribbon" connecting the two surfaces:



In other words,

$$\Delta\Phi_B = \int_{\text{ribbon}} \vec{B} \cdot d\vec{a} \, .$$

Now, we want to think about how we actually do this integral. Before setting it up, let's think for a moment about charges that might be moving in the wire loop. Let's suppose a charge's velocity along the loop is $\vec{u}$. The *total* velocity of the change is then $\vec{v}_{\text{charge}} = \vec{u} + \vec{v}$ — we add the velocity along the loop to the "external" velocity of the loop.

OK, now think about the area element on the ribbon. By staring at the figures a little bit and thinking about what the area element means, we can see that

$$d\vec{a} = (\vec{v}\,\Delta t) \times d\vec{l} \, .$$

In fact, since $\vec{u}$ is parallel to $d\vec{l}$, we can just as well write this as

$$d\vec{a} = (\vec{v}_{\text{charge}}\Delta t) \times d\vec{l}$$

— this is *exactly* the same as the previous way of writing $d\vec{a}$ since $\vec{u} \times d\vec{l} = 0$.

We have thus reduced our *area* integral to a *line* integral: we just have to integrate with respect to $d\vec{l}$ around the wire loop. In other words,

$$\Delta\Phi_B = \oint_{\text{loop}} \vec{B} \cdot \left[(\vec{v}_{\text{charge}}\Delta t) \times d\vec{l}\right] \, .$$

Since we are holding Δt fixed while we do the integral, we can divide it out and take the limit $\Delta t \to 0$:

$$\frac{\partial\Phi_B}{\partial t} = \oint_{\text{loop}} \vec{B} \cdot \left(\vec{v}_{\text{charge}} \times d\vec{l}\right) \, .$$

Almost there! To massage this further, note that, for any three vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$,

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = (\vec{a} \times \vec{b}) \cdot \vec{c} \, .$$

(This is very simple to prove. The most straightforward way is to just write out $\vec{a} = a_x\hat{x} + a_y\hat{y} + a_z\hat{z}$, etc, and grind. If you know about the properties of matrix determinants, you can probably come up with an even quicker proof.) We then rearrange our integrand to give us

$$\begin{aligned}
\frac{\partial \Phi_B}{\partial t} &= \oint_{\rm loop} \left(\vec{B} \times \vec{v}_{\rm charge}\right) \cdot d\vec{l} \\
&= -\oint_{\rm loop} \left(\vec{v}_{\rm charge} \times \vec{B}\right) \cdot d\vec{l} \\
&= -c\oint_{\rm loop} \left(\frac{\vec{v}_{\rm charge}}{c} \times \vec{B}\right) \cdot d\vec{l}
\end{aligned}$$

The quantity in parentheses on the last line is just the magnetic force (per unit charge). By integrating it over a closed loop, we are computing the work done, per unit charge, to take it around the loop. By definition, this is the EMF! We thus have finally proven our general result:

$$\frac{\partial \Phi_B}{\partial t} = -c\mathcal{E}$$

or

$$\mathcal{E} = -\frac{1}{c}\frac{\partial \Phi_B}{\partial t}\ .$$

14.5 A few subtleties

14.5.1 Work

In the above discussion, I claimed that $(\vec{v}_{\rm charge}/c) \times \vec{B}$, integrated over a closed loop, is the work done per unit charge to go around the loop; this is how I defined the EMF $\mathcal{E}$ we just calculated. This hopefully bothered the hell out of you: a few lectures ago, I claimed that magnetic forces never do work on charges!

Am I liar? Not in this instance; instead, there is something subtle going on. Work is being done — but it is not actually being done by the magnetic field! Instead, the work is being done by whatever "external agent" moves the loop. The loop could be moving because it is being dragged by a team of sled dogs; or, it could be part of an electric generator; or it could be falling in a gravitational field. In cases like this, one can look more carefully at the work that is being done by this "agent" and show that it is precisely what is needed to get the EMF I claimed above.

14.5.2 Holding loop fixed, moving the field

In the above discussion, we have fixed our magnetic field and moved our loop of wire. What if instead we held our loop fixed and moved the magnetic field? (Or rather, the source of the field?)

From our discussion of special relativity, you should be able to guess the answer: we had best get the same result! Moving the magnetic field and moving the loop have to be exactly equivalent to one another — we're just jumping from one frame to another.

However, when the loop is held steady, the charges don't move. There is thus no way for a magnetic field to exert a force on them. The only way to reconcile the current that we see induced in the loop is to conclude that there must be an electric field present.

This electric field *cannot* be a "nice" electric field, like we got to know back when we only worried about electrostatics. It is an electric field that drives current to run around in a circle. From the idea that the EMF is the work per unit charge, it follows that

$$\mathcal{E} = \oint_C \vec{E}_{\rm ind} \cdot d\vec{l}$$

where $\vec{E}_{\rm ind}$ is the electric field that is *induced* by the changing magnetic field, and C is the closed path around the loop. But we also have

$$\mathcal{E} = -\frac{1}{c}\frac{\partial \Phi_B}{\partial t} \ .$$

Let's put these two equations together:

$$\begin{aligned} \oint_C \vec{E} \cdot d\vec{l} &= -\frac{1}{c}\frac{\partial \Phi_B}{\partial t} \\ &= -\frac{1}{c}\frac{\partial}{\partial t}\int_S \vec{B} \cdot d\vec{a} \ . \end{aligned}$$

The surface S is any surface which has the closed path C as its boundary. (We drop the subscript "ind" on the electric field from now on; it should be clear that we are discussing an electric field that arises from the variations of the magnetic field.)

The logic should smell kind of familiar at this point. We've got a line integral on the left, and a surface integral on the right. To put things into a "nice" form, we want to change our line integral into a surface integral. We thus invoke Stokes' theorem:

$$\oint_C \vec{E} \cdot d\vec{l} = \int_S (\vec{\nabla} \times \vec{E}) \cdot d\vec{a} \ .$$

Plugging this in, we find

$$\begin{aligned} \int_S \left(\vec{\nabla} \times \vec{E}\right) \cdot d\vec{a} &= -\frac{1}{c}\frac{\partial}{\partial t}\int_S \vec{B} \cdot d\vec{a} \\ \int_S \left(\vec{\nabla} \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t}\right) \cdot d\vec{a} &= 0 \end{aligned}$$

Since this must hold for an arbitrary surface S bounded by the curve C, we conclude that

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \ .$$

This equation quantitatively relates the varying magnetic field to the induced electric field.

14.6 Faraday's law

The most important equations that we have derived in this lecture are all different variants of what is called **Faraday's law**:

$$\mathcal{E} = \oint_C \vec{E} \cdot d\vec{l} = -\frac{1}{c}\frac{\partial \Phi_B}{\partial t} = -\frac{1}{c}\frac{\partial}{\partial t}\int_S \vec{B} \cdot d\vec{a}$$

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\frac{\partial B}{\partial t} \; .$$

The first version is the "integral form" of Faraday's law; the second is the "differential form".

It's worth stopping at this point and gathering together all of the equations we've got so far describing divergences and curls of electric and magnetic fields:

$$\vec{\nabla} \cdot \vec{E} = 4\pi\rho$$

$$\vec{\nabla} \cdot \vec{B} = 0$$

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

$$\vec{\nabla} \times \vec{B} = \frac{4\pi}{c}\vec{J} \; .$$

These are *almost* the Maxwell's equations that completely describe the electromagnetic field. There's just one tiny little piece missing. Maxwell figured out what this missing piece was just by arguing that the two curl equations should exhibit a kind of symmetry which is lacking in their current form. You may be able to figure out what's missing by taking the divergence of the last equation and noticing that there's an inconsistency. Once you fix up this inconsistency, you get a set of equations suitable for printing on a t-shirt.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 15:
MUTUAL AND SELF INDUCTANCE.

15.1 Using induction

Induction is a fantastic way to create EMF; indeed, almost all electric power generation in the world takes advantage of Faraday's law to produce EMF and drive currents. Given the defining formula,

$$\mathcal{E} = -\frac{1}{c}\frac{d\Phi_B}{dt}$$

where

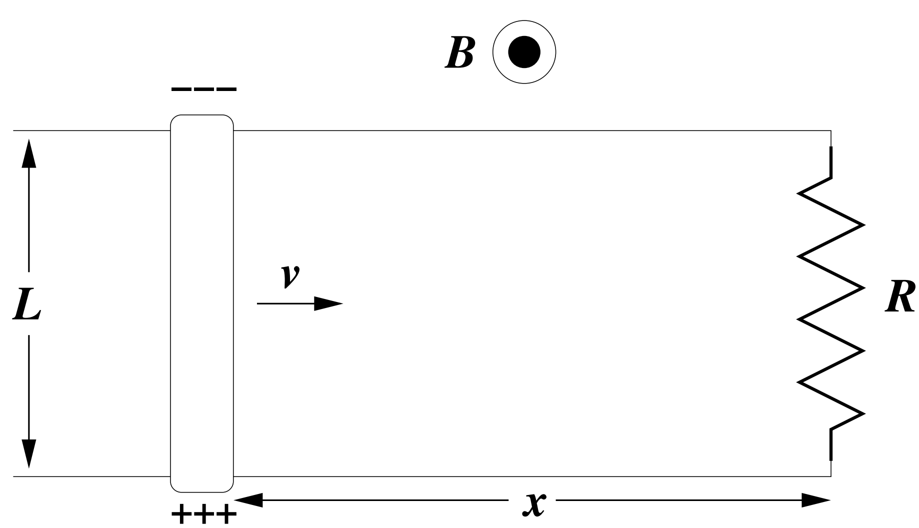
$$\Phi_B = \int_S \vec{B} \cdot d\vec{A}$$

all that we need to do is set up some way to make a time varying magnetic flux and we can induce EMF and drive currents very happily.

Given the definition of magnetic flux, there are essentially three ways that we can make it vary and thereby create an EMF: we can make the area vary; we can make the "dot product" vary; and we can make the magnetic field vary. Let's look at these three situations one by one.

15.1.1 Changing area

We already discussed exactly this situation when we looked at the rod sliding on rails in a uniform magnetic field:

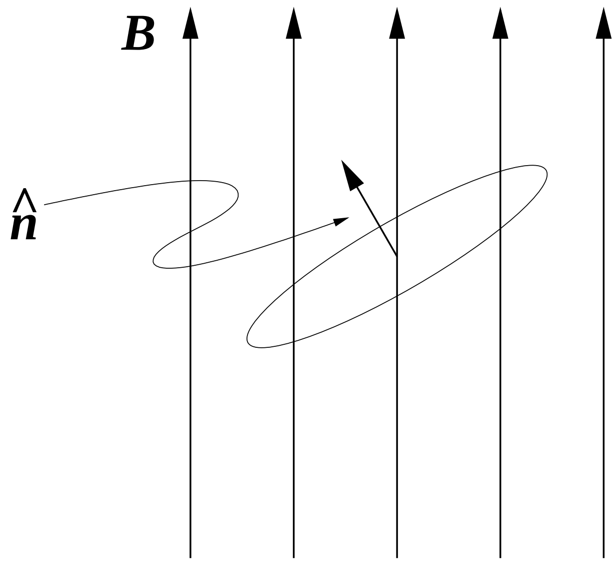


The changing flux in this case is due to the continually changing area. We get an EMF whose magnitude is $\mathcal{E} = BLv/c$, and that (by Lenz's law) drives a current which circulates counterclockwise in the loop.

In principle, we could imagine using this as a power source. For example, we could replace the resistor with a lightbulb and use this arrangement as a lamp. However, you'd have to find some way to continually slide the bar: the energy radiated by the lamp has to come from somewhere! In this case, it would come from the poor schmuck who continually drags the bar back and forth.

15.1.2 Changing "dot product"

Another way to generate an EMF is to make the "dot product" — i.e., the relative orientation of our area and the magnetic field — vary with time. Imagine we have a circular loop sitting in a uniform magnetic field:



The magnetic flux through the loop is given by

$$\begin{aligned}\Phi_B &= \pi r^2 \vec{B} \cdot \hat{n} \\ &= \pi r^2 B \cos\theta\end{aligned}$$

where θ is the angle between the loop's normal vector and the magnetic field. Now, suppose we make the loop rotate: say the angle varies with time, $\theta = \omega t$. The EMF that is generated in this case is then given by

$$\mathcal{E} = \frac{\omega \pi r^2 B}{c} \sin\omega t\ .$$

This kind of arrangement is in fact exactly how most power generators operate: An external force (flowing water, steam, gasoline powered engine) forces loops to spin in some magnetic field, generating electricity. Note that the EMF that is thereby generated is sinusoidal — generators typically produce AC (alternating current) power.

15.1.3 Changing field

Finally, in many situations we can make the magnetic field itself vary. Suppose the integral is very simple, so that it just boils down to field times area. If the magnetic field is a function of time, then the flux will be a function of time:

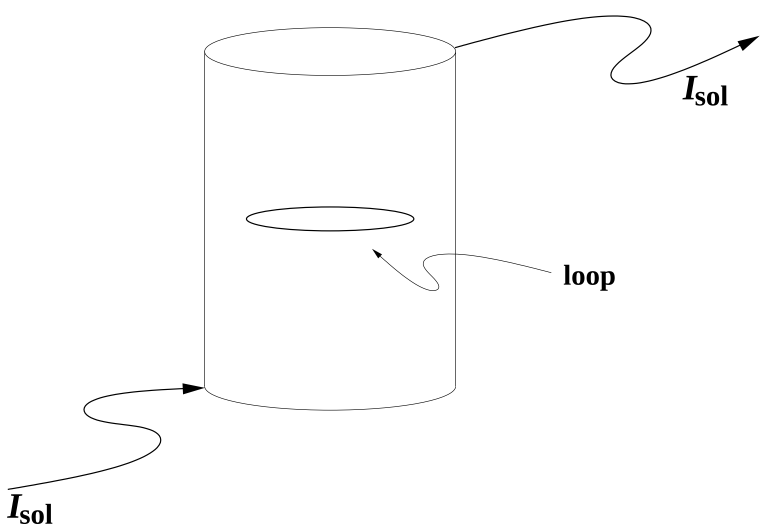
$$\Phi_B(t) = B(t)A \ .$$

We just need to take a derivative, divide by c, and we have an EMF.

This method of creating an EMF is one of the most important. In practice, it typically comes about when our magnetic field comes from some current that varies in time. This is the case that we will focus on in the remainder of this lecture.

15.2 Inductance example

Consider the following arrangement: we place a loop of wire with radius r in the center of a long solenoid. The solenoid has a wire that is wrapped N times around its total length l; hence, the number of turns per unit length is N/l. A current $I_{\rm sol}$ runs through this solenoid; this current varies with time. What is the EMF $\mathcal{E}_{\rm loop}$ that is induced in the wire loop?



We begin with our defining relation,

$$\mathcal{E} = \frac{1}{c}\frac{d\Phi_B}{dt} \ .$$

(In what follows, I will often drop the minus sign. This is fine provided we bear in mind that we need to use Lenz's law to tell us the direction in which any induced current will flow.) The magnetic flux is given by the magnetic field of the solenoid times the cross sectional area of the loop:

$$\begin{aligned}\Phi_B &= B_{\rm sol}A_{\rm loop} \\ &= \frac{4\pi N I_{\rm sol}}{cl}\pi r^2 \ .\end{aligned}$$

(Recall that the current per unit length is nI if the input current is I and the number of turns per unit length is n.) This means that the EMF induced in the little loop is

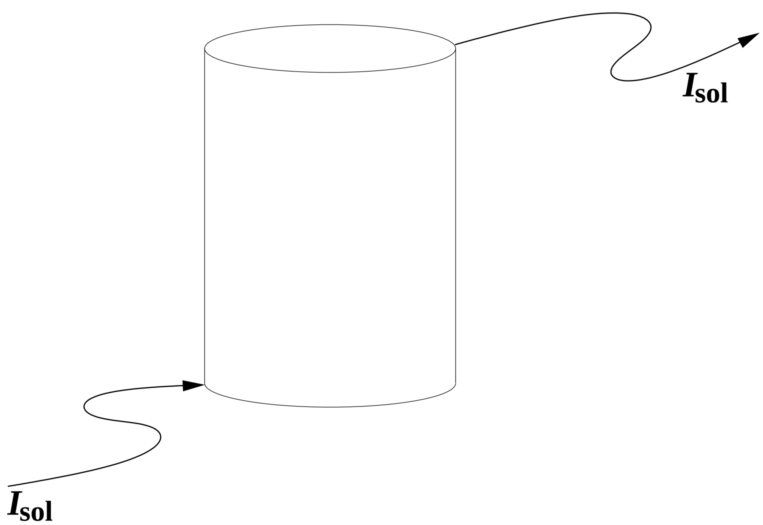
$$\mathcal{E}_{\rm loop} = \frac{4\pi^2 r^2 N}{c^2 l}\frac{dI_{\rm sol}}{dt}\ .$$

In words, the EMF induced in the loop is totally set by the rate of change of current in the solenoid, times a bunch of constants that depend on the geometry of our setup.

We will return to look at this example somewhat more systematically.

15.3 Self inductance

Suppose we take the loop out of the solenoid:



Is there any EMF induced in this case? Let's think about this carefully: The magnetic field threading the solenoid is

$$B = \frac{4\pi N I_{\rm sol}}{cl}\ .$$

The magnetic flux through *each loop* that constitutes the solenoid is

$$\Phi_B^{1\ \rm loop} = \frac{4\pi^2 R^2 N I_{\rm sol}}{cl}$$

where R is the solenoid's radius. This means that the *total* flux as far as any induction is concerned must be $\Phi_B^{\rm tot} = N\Phi_B^{1\ \rm loop}$ — we get a contribution of $\Phi_B^{1\ \rm loop}$ from every loop that constitutes the solenoid! In other words, we end up with

$$\Phi_B^{\rm tot} = \frac{4\pi^2 R^2 N^2 I_{\rm sol}}{cl}\ .$$

Taking the time derivative, we find the EMF that is induced when this current varies:

$$\mathcal{E} = \frac{4\pi^2 R^2 N^2}{c^2 l}\frac{dI_{\rm sol}}{dt}\ .$$

Again, we have a bunch of stuff that just depends on the solenoid's geometry (and constants) times the rate of change of the current.

The math hopefully makes sense. What you should be asking yourself at this point is "OK, fine, I've got an EMF, but what *IS* it??" Lenz's law helps us to understand what this EMF is and does: Suppose the current is increasing, so that the magnetic field is likewise increasing. Lenz's law tells us that the induced EMF will try to "fight" this: the EMF that is induced will tend to drive currents which *oppose* the change in magnetic flux. For this setup, this means that the EMF will oppose the increase — it will drive a current which tries to prevent B from changing. If the current was decreasing, the EMF will oppose the decrease. The punchline is that *inductance opposes the change in the current.* The EMF that is induced in a situation like this is often called the "back EMF" since it acts "back" on the circuit to try to hold the current steady.

From this situation, we define a quantity called the "self-inductance", L:

$$\mathcal{E} = L\frac{dI}{dt}\ .$$

Quantities which have a self-inductance and are used in circuits (as we'll study next lecture) are called "inductors". For a solenoidal inductor, we can read off the value of L from the preceding calculation:

$$L_{\rm sol} = \frac{4\pi^2 N^2 R^2}{c^2 l}\ .$$

Bear in mind that this formula *only* holds for solenoids! The solenoid is, however, a very important example of an inductor and crops up quite a lot.

15.3.1 Units

Since EMF has units of potential, and dI/dt is [charge]/[time]^2, it is simple to deduce the units of self-inductance. In cgs units, we have

$$\frac{[\text{Potential}]}{[\text{charge}]/[\text{time}]^2} \to \frac{\text{esu/cm}}{\text{esu/sec}^2} \to \frac{\text{sec}^2}{\text{cm}}\ .$$

Inductance is measured in the rather odd unit of sec^2/cm in cgs units.

In SI units, we have

$$\frac{[\text{Potential}]}{[\text{charge}]/[\text{time}]^2} \to \frac{\text{Volts}}{\text{Amp/sec}}\ .$$

We could reduce this further, but what is done instead is to give this combination a name: 1 Volt/(Amp sec^{-1}) = 1 Henry. This unit is named after Joseph Henry, who discovered self-inductance. There's no question it's a pretty goofy name for a unit, though.

15.4 Energy stored in an inductor

Suppose we have an inductor that is sitting on its own, and we somehow force a current to flow through it. As soon as the current level starts changing, a back EMF is induced that opposes the flow of the current. We can *force* the current to flow by fighting this back EMF;

but, it takes power to do so. To overcome the back EMF of the inductor, we must supply a power to the inductor of

$$P = I\mathcal{E} = LI\,\frac{dI}{dt}\ .$$

Suppose we increase the current flowing through the inductor from $I = 0$ at $t = 0$ to I at some later time t. How much work is done to do this? Since $P = dW/dt$, we just integrate:

$$\begin{aligned} W &= \int_0^t P\,dt \\ &= \int_0^t LI\,\frac{dI}{dt}\,dt \\ &= \int_0^I LI\,dI \\ &= \frac{1}{2}LI^2\ . \end{aligned}$$

Since we did this much work to get the current flowing, this is the energy that is stored in the inductor:

$$U_L = \frac{1}{2}LI^2\ .$$

15.4.1 Magnetic field energy

How exactly is this energy stored in the inductor? The key thing that we changed is that we created a magnetic field where there was none previously. This energy must essentially be the energy that is stored in the magnetic field we have just created! (Since we previously talked about energy stored in electric fields, and since electric and magnetic fields are essentially the same thing from the viewpoint of special relativity, it shouldn't be too much of a surprise that we can discuss the energy of magnetic fields as well.)

It is easy to see that a magnetic field has an energy density by looking at the specific example of the solenoidal inductor. Suppose a current I is flowing in the solenoid discussed previously. The energy that is stored in it is $U_L = \frac{1}{2}LI^2$. We also know that $L = 4\pi^2N^2R^2/(c^2l)$. Let's write this out and reorganize things:

$$\begin{aligned} U_L &= \frac{1}{2}LI^2 \\ &= \frac{2\pi^2N^2R^2}{c^2l}I^2\ . \end{aligned}$$

The volume of the solenoid is just $V = \pi R^2 l$ (it's a cylinder of radius R and height l). The magnetic field of the solenoid is $B = 4\pi NI/(cl)$. Knowing this, we can rewrite the above formula as

$$\begin{aligned} U_L &= \frac{\pi R^2 l}{8\pi}\left(\frac{4\pi NI}{cl}\right)^2 \\ &= (\text{Volume})\,\frac{B^2}{8\pi}\ . \end{aligned}$$

This means that the magnetic field has an energy per unit volume of $B^2/8\pi$ — a result that is *VERY* similar to the energy per unit volume of the electric field!

This is one of the nice things about cgs units: it really brings out the symmetry between the fields. If u_X is the energy per unit volume of the X field, then we have

$$\begin{aligned} u_E &= \frac{E^2}{8\pi} \\ u_B &= \frac{B^2}{8\pi}\,. \end{aligned}$$

In SI units, it's not quite so symmetric:

$$\begin{aligned} u_E &= \frac{\epsilon_0\, E^2}{2} \\ u_B &= \frac{B^2}{2\mu_0}\,. \end{aligned}$$

15.4.2 Using energy to find L

In many cases, the result that the energy density $u_B = B^2/8\pi$ and the total energy $U = \frac{1}{2}LI^2$ can be used as a tool for computing the self inductance. This is sometimes easier than looking at a time varying current, calculating the EMF induced, and reading out the inductance L. Here's a simple recipe that often works very nicely:

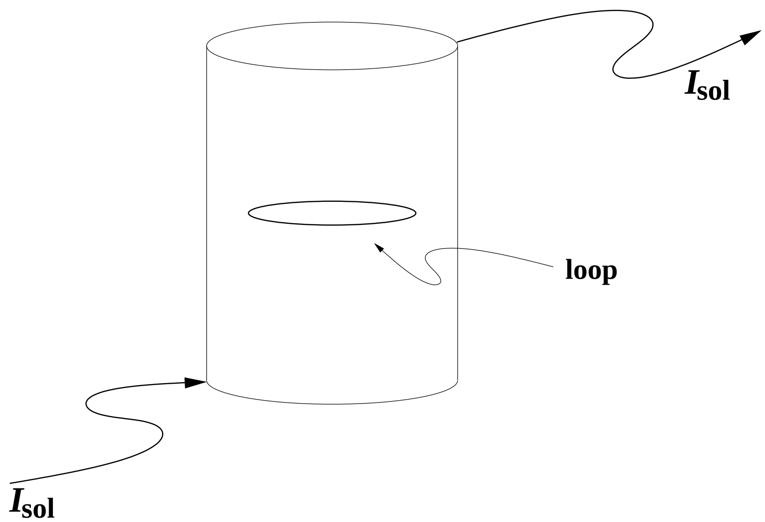
- Compute the magnetic field.
- Integrate $B^2/8\pi$ over a volume.
- Divide by I^2; multiply by 2.

What is left must be the inductance L.

You will have an opportunity to test drive this trick on a pset.

15.5 Mutual inductance and reciprocity

Let's return to the first example that we discussed: the loop inside the solenoid. Let us label the solenoid with a "1" (i.e., it is "circuit element 1"), and the loop with a "2". We



found that the EMF induced in the loop was equal to some factor times the rate of change of current in the solenoid. Using our new labeling system, we would say

$$\mathcal{E}_2 = M_{21}\frac{dI_1}{dt} \; .$$

The coefficient M_{21} is called the *mutual inductance*; in this case it is the mutual inductance from element 1 to element 2. For this setup, it is something we already worked out:

$$M_{21} = \frac{4\pi^2 r^2 N}{c^2 l}$$

where r is the radius of the small loop.

Suppose that we now reverse things: rather than putting our time varying current through the solenoid, we put it through the small loop. By definition, the EMF induced in the solenoid by this time changing current must be described by

$$\mathcal{E}_1 = M_{12}\frac{dI_2}{dt} \; .$$

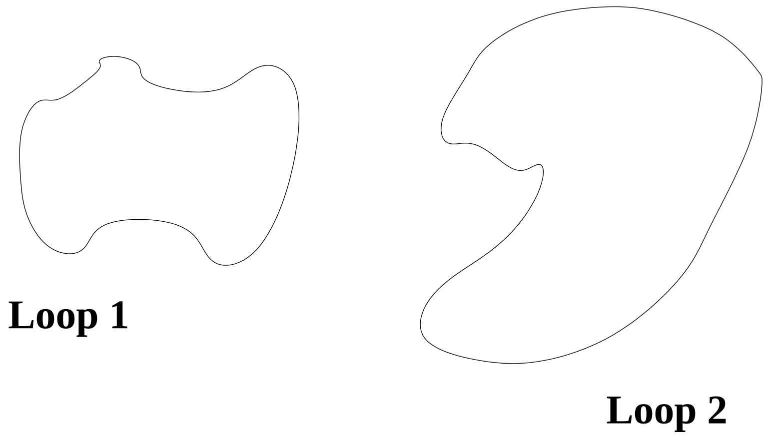
However, *how* do we calculate M_{12}? It should be a total mess — the B field from the loop is *highly* non trivial, and the geometry of the solenoid through which we need to calculate the flux is very complicated. It seems like there is *no hope* to possibly calculate this!

It turns out there is an amazing theorem — the "reciprocity theorem" — that makes it all easy. The reciprocity theorem tells us that

$$M_{12} = M_{21} \; .$$

Operationally, this means we can solve the "hard" problem — induced EMF in solenoid due to changing currents in loop — by first solving the "easy" problem — induced EMF in loop due to changing currents in solenoid. This is fantastic news for solving nasty problems.

The proof of this result is a very nice application of the vector potential. Consider two loops of wire,



Suppose that a current I runs through loop 1. What is the flux of magnetic field through loop 2 due to the current in loop 1? By definition, this flux will be

$$\Phi_{21} = \int_{S_2} \vec{B}_1 \cdot d\vec{a}_2$$

where $\vec{B}_1$ is the magnetic field arising from loop 1, S_2 is the surface spanned by loop 2, and $d\vec{a}_2$ is an area element on this surface. (You'll see why I choose a lower case a for the area element in just a moment.) Now, we rewrite this result in terms of the vector potential arising from loop 1:

$$\begin{aligned}\Phi_{21} &= \int_{S_2} \left(\vec{\nabla} \times \vec{A}_1\right) \cdot d\vec{a}_2 \\ &= \oint_{C_2} \vec{A}_1 \cdot d\vec{l}_2 \ .\end{aligned}$$

On the last line, we've used Stokes' theorem to reexpress the flux as a line integral of the vector potential. Finally, we use the fact that the vector potential from a current in loop 1 to some field point is

$$\vec{A}_1 = \frac{I}{c} \oint_{C_1} \frac{d\vec{l}_1}{r} \ .$$

With this, the flux becomes

$$\Phi_{21} = \frac{I}{c} \oint_{C_1} \oint_{C_2} \frac{d\vec{l}_1 \cdot d\vec{l}_2}{r_{21}}$$

where r_{21} is the distance from the length element $d\vec{l}_1$ to the length element $d\vec{l}_2$.

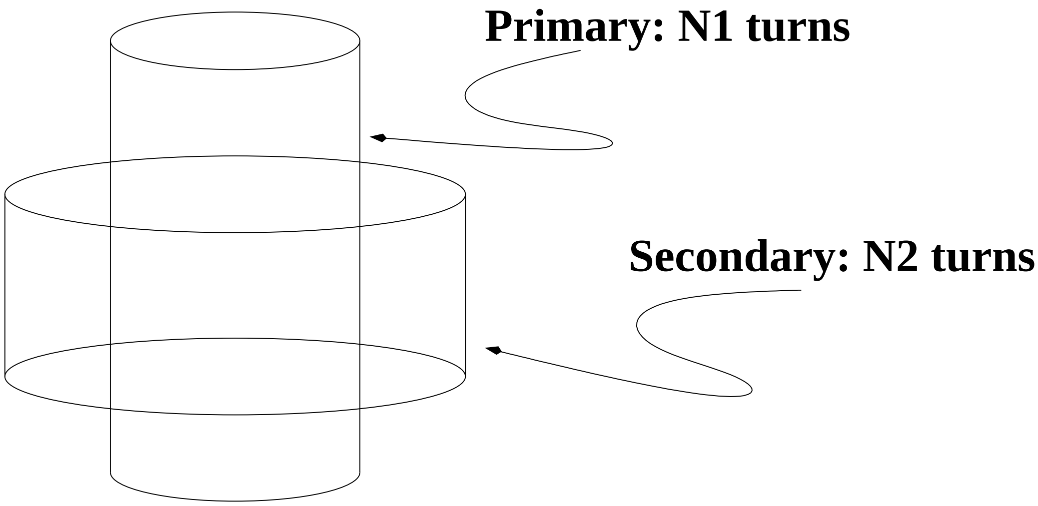
Now, what happens if I put the current I in loop 2 and ask what is the flux that goes through loop 1? Going through all of these steps again, we find

$$\Phi_{12} = \frac{I}{c} \oint_{C_2} \oint_{C_1} \frac{d\vec{l}_2 \cdot d\vec{l}_1}{r_{12}} \ .$$

This is the exact same thing as Φ_{21}! Since the fluxes are the same, the derivative of the fluxes must be the same; it follows that the mutual inductance $M_{12} = M_{21}$, and the reciprocity theorem is proved.

15.6 Transformers

A very important *practical* application of mutual inductance is in the construction of transformers: devices which step up (or step down) an AC voltage. The simplest transformer consists of a pair of nested solenoids:



The idea is that we drive a time varying current through the primary, making a time varying magnetic flux:

$$\mathcal{E}_1 = \frac{N_1}{c}\frac{d\Phi_B}{dt}$$

where Φ_B is the magnetic flux from a single turn. Since the same flux threads the second solenoid, the induced EMF in the outer solenoid is

$$\mathcal{E}_2 = \frac{N_2}{c}\frac{d\Phi_B}{dt} \ .$$

Comparing, we have

$$\mathcal{E}_2 = \mathcal{E}_1 \frac{N_2}{N_1} \ .$$

By changing the ratio of turns, we can make enormous voltages ... or, we can step voltages down.

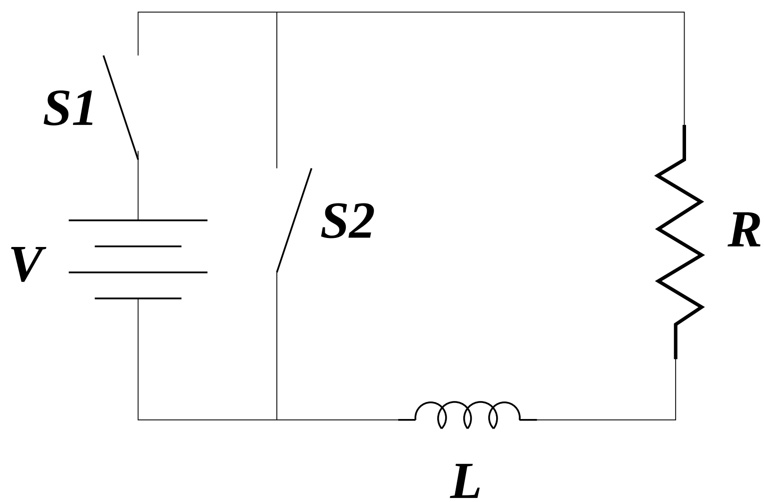
Scott Hughes *12 April 2005*

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 16:
RL Circuits.
Undriven RLC circuits; phasor representation.

16.1 RL circuits

Now that we know all about inductance, it's time to start thinking about how to use it in circuits. A simple circuit that illustrates the major concepts of inductive circuits is this:



Suppose that at $t = 0$, we close the switch $S1$, leaving $S2$ open. How does the current evolve in the circuit after this?

Let's first think about this physically. Lenz's law tells us that the magnetic flux through the inductor does not "want" to change. Since this flux is initially zero, the inductor will impede any current that tries to flow it — the current will initially try to remain at zero. As time passes, current will gradually leak through, and the magnetic flux will build up. Eventually, it should saturate at $I = V/R$ — at this point the most current possible is flowing through the circuit.

16.1.1 To Kirchhoff or not to Kirchhoff

We'd now like to analyze this circuit quantitively as well. In all other circumstances in which we've examined a circuit, we've used Kirchhoff's laws for this; for a a simple single-loop circuit, the relevant rule is Kirchhoff's second law, "The sum of the EMFs and the voltage drops around a closed loops is zero".

This rule is essentially a restatement of the old electrostatics rule that $\oint \vec{E} \cdot d\vec{s} = 0$. **This rule DOES NOT HOLD when we have a time changing magnetic field!** In other words, inductors completely invalidate — at least formally — the foundation of Kirchhoff's second law.

So what do we do? Let's reconsider: in place of $\oint \vec{E}\cdot d\vec{s} = 0$, we have Faraday's law:

$$\oint \vec{E}\cdot d\vec{s} = -\frac{1}{c}\frac{d\Phi_B}{dt} = \mathcal{E}_{\rm ind} = L\frac{dI}{dt}\ .$$

The closed loop integral of $\vec{E}$ gives us the induced EMF. I personally find it best to think of the inductor as an EMF source; it acts in the circuit effectively like a battery would. Even though Kirchhoff's laws no longer stand, formally, on a very solid foundation, we can extend the formalism by thinking of inductors in the circuit as a source of EMF. The equations then carry over just fine; indeed, they carry over so well that almost all textbooks tell you to just apply Kirchhoff's laws to circuits containing inductors without even mentioning this subtlety. It's always worth bearing in mind that what you are *really* using is Faraday's law. I will try to refer to the loop rule as Faraday's law; it's such a habit to call it Kirchhoff's law, though, that I will surely slip from time to time.

The tricky thing in these kinds of problems is getting the signs right. In *all* of these cases, Lenz's law should guide you: the induced EMF will act in such a way as to oppose changes in the inductor's magnetic flux. If a circuit initially has no current flowing through it, the inductor will oppose the build-up of current; if it initially *does* have current flowing, the inductor will try to keep that current flowing. Keep this intuition in mind and you should have *no* problem choosing the correct sign for the inductor's EMF.

(Very detailed discussion of inductors and the failure of Kirchhoff's laws can be found in Walter Lewin's OCW 8.02 lectures from 2002; see Lectures 15 and 20 in particular.)

16.1.2 Back to our regularly scheduled program

OK, let's calculate. The EMF due to the battery is $+V$; the EMF induced in the inductor is $-L dI/dt$. Why the minus sign? It's going to fight the growth in current, and so must be opposite to the battery. Finally, there's a voltage drop $-IR$ at the resistor:

$$V - IR - L\frac{dI}{dt} = 0\ .$$

Rearrange:

$$\begin{aligned}\frac{dI}{dt} &= \frac{V-IR}{L}\\ &= \frac{R}{L}\left(\frac{V}{R} - I\right)\ .\end{aligned}$$

Now divide; we end up with something that is of the form du/u:

$$\frac{dI}{I - V/R} = -\frac{R\,dt}{L}$$

so it is easy to integrate up, using the boundary conditions $I = 0$ at $t = 0$, $I = I(t)$ at some general time t:

$$\ln\left[\frac{I(t) - V/R}{-V/R}\right] = -\frac{Rt}{L}$$

or

$$I(t) = \frac{V}{R}\left(1 - e^{-Rt/L}\right)\ .$$

As we guessed on physical grounds, the current builds up, with a "growth timescale" L/R. The timescale L/R plays a role in LR circuits similiar to that played by the timescale RC in circuits with capacitors.

Suppose we let this circuit sit in this state for a very long time, so that the current saturates at its equilibrium value, $I = V/R$. What happens if we now simultaneously open $S1$ and close $S2$? Physically, our expectation is that Lenz's law will try to hold the current flow constant — the inductor says "I don't want the flux through me to change!" However, we know that the resistor dissipates energy, so the current must bleed away — we should see some kind of exponential decay.

Time to calculate. Going around the circuit, we have

$$-L\frac{dI}{dt} - IR = 0 .$$

The first term is the EMF due to the inductor, $-LdI/dt$. Why the minus sign? By Lenz's law, the inductor always fights change. If the current "wants" to be reduced, the sign of the inductor's EMF needs to fight that. The second term is the usual voltage drop across a resistor. Rearranging leads to the easily integrated equation

$$\frac{dI}{I} = -\frac{R\,dt}{L} ,$$

which in turn leads to the exponential solution

$$I(t) = I(t = 0)e^{-Rt/L} .$$

(Note that we have redefined the meaning of $t = 0$ — this is now the time at which we open $S1$ and close $S2$.)

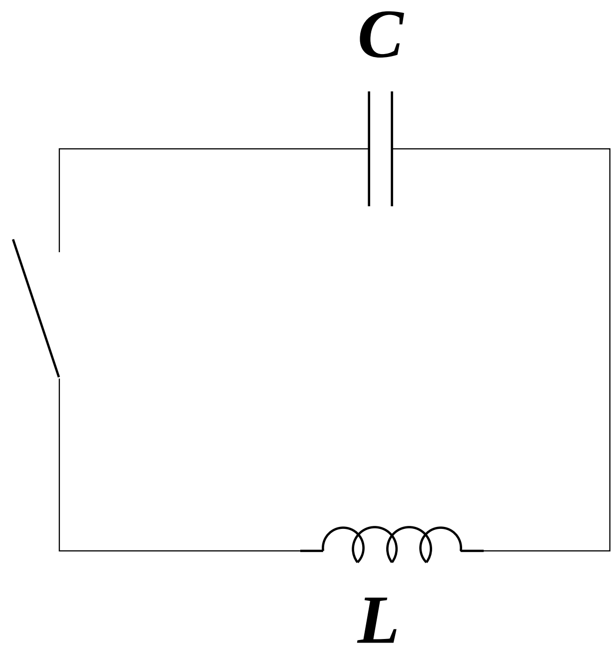
When thinking about *any* circuit with inductance, the key thing you should keep in mind is that inductors cause currents to have a kind of "inertia". If there is no current flowing, the inductor will force the current to build up gradually; if there is current flowing, the inductor will force it to change gradually. If current is already flowing and there is a sudden break in the circuit, the inductor will do whatever it takes to keep the current flowing. A consequence of this is that when we open switches or break circuits very suddenly, we may see a large spark! That spark is because of Lenz's law — inductors in the circuit refuse to let the flux through them change. If that means that they must force the current to flow through the air, creating a huge, fiery spark — so be it.

16.1.3 Unit check

Before moving on, let's do the traditional sanity check on our units: the ratio L/R should have the dimension of time for any of the above results to make sense. In cgs, the units of L are $\mathrm{sec}^2/\mathrm{cm}$; the units of R are sec/cm. So L/R is indeed a time measured in seconds. In SI, the unit of L is the Henry, which is Volt-sec/Amp; the unit of R is the Ohm, which is Volt/Amp. So L/R is again a time measured in seconds. Three cheers for dimensional analysis.

16.2 LC circuits

Exponential decay (or saturation) is dull. We're now going to look at a circuit configuration that does something new. Consider what happens when we connect an inductor to a capacitor:



Suppose that the capacitor is charged up to a charge separation Q_0. We close the switch at $t = 0$. How does the circuit behave after this?

As usual, our governing analysis of the system comes from Kirchhoff/Faraday: going around the circuit, we have

$$\frac{Q}{C} - L\frac{dI}{dt} = 0$$

where Q is the charge on the capacitor, and I is the current flowing through the inductor. To get anywhere with this, we of course need to relate the current flowing to the charge on the capacitor. Since the current comes from charge flowing off the capacitor, we have

$$I = -\frac{dQ}{dt}\ .$$

We thus finally end up with the differential equation

$$\frac{d^2Q}{dt^2} + \frac{1}{LC}Q = 0\ .$$

Differential equations of this form are best solved via an age old technique: guessing. In particular, guided by our experience with other systems that have a similar form, we guess that the solution has the form

$$Q(t) = A\cos\omega_0 t + B\sin\omega_0 t\ .$$

We plug this solution back into our differential equation; the key thing we need is

$$\begin{aligned}\frac{d^2Q}{dt^2} &= -\omega_0^2 A\cos\omega_0 t - \omega_0^2 B\sin\omega_0 t\\ &= -\omega_0^2 Q\ .\end{aligned}$$

Our assumed solution works fine provided we set the angular frequency

$$\omega_0 = \frac{1}{\sqrt{LC}} \; .$$

(You should be able to verify immediately that this has the correct units.) The constants A and B are free to be adjusted to match our initial conditions — the value of the charge on the capacitor Q at $t = 0$, and the current flowing in the circuit at $t = 0$:

$$\begin{aligned} Q(0) &= Q_0 \\ &= A\cos(0) + B\sin(0) \\ \rightarrow A &= Q_0 \; . \end{aligned}$$

$$\begin{aligned} I(0) &= -\left(\frac{dQ}{dt}\right)_{t=0} = 0 \\ &= -\omega_0 A\sin(0) + \omega_0 B\cos(0) \\ \rightarrow B &= 0 \; . \end{aligned}$$

We now have the full solution to this problem: the charge on the capacitor, as a function of time, is

$$Q(t) = Q_0 \cos\omega_0 t \; .$$

Equivalently, the current flowing in the circuit, $I(t) = -dQ/dt$, is

$$\begin{aligned} I(t) &= \omega_0 Q_0 \sin\omega_0 t \\ &= \frac{Q_0}{\sqrt{LC}} \sin\omega_0 t \; . \end{aligned}$$

This sinusoidal behavior is the generic feature of LC circuits. Notice that the charge has maximum magnitude when the current is zero, and vice versa — the current and the charge are "90° out of phase" with each other.

A nice way to make sense of the current and charge behavior is by looking at the energetics of the circuit elements. The energy stored in the capacitor, as a function of time, is

$$\begin{aligned} U_C(t) &= \frac{Q(t)^2}{2C} \\ &= \frac{Q_0^2}{2C}\cos^2\omega_0 t \; . \end{aligned}$$

The energy stored in the inductor, as a function of time, is

$$\begin{aligned} U_L(t) &= \frac{1}{2} L I(t)^2 \\ &= \frac{1}{2}\frac{1}{LC} L Q_0^2 \sin^2\omega_0 t \\ &= \frac{Q_0^2}{2C}\sin^2\omega_0 t \; . \end{aligned}$$

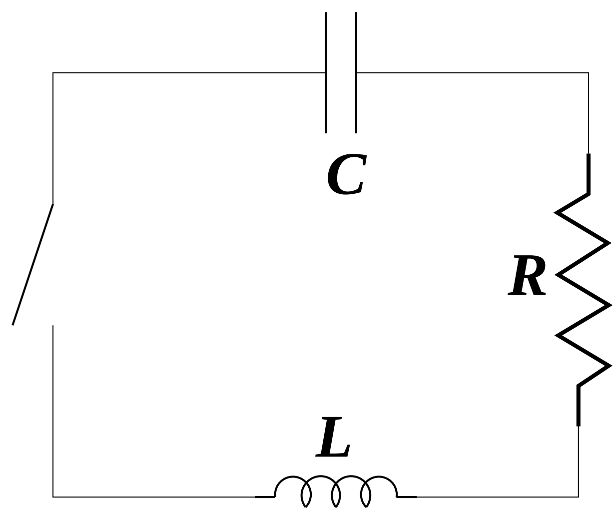
The energy in the inductor and the energy in the capacitor each oscillate. Note that their sum is constant, and is exactly equal to the total energy that was initially stored in the capacitor:

$$\begin{aligned} U_{\rm tot} &= U_C(t) + U_L(t) = \frac{Q_0^2}{2C}\left(\cos^2\omega_0 t + \sin^2\omega_0 t\right) \\ &= \frac{Q_0^2}{2C}\,. \end{aligned}$$

Conservation of energy works! In this case, there's an even cooler interpretation. The energy in the capacitor is stored as electric fields; that in the inductor is stored as magnetic fields. What we are seeing here is that the system's total energy is just sloshing back and forth, from electric to magnetic to electric to ... This circuit is just an electromagnetic swing set!

16.3 RLC circuits

The circuit we just discussed is an interesting one ... but it's also somewhat fictional, since we are idealizing it as having no resistance. Since *all* circuits have *some* resistance[1], the RLC circuit sketched here is *far* more relevant to real world applications than the LC circuit:



How would we expect this circuit to behave? Our intuition on the energetics of LC circuits is helpful here: qualitatively, we should expect the same "sloshing" of energy between the inductor and capacitor. However, we now expect that some energy will be dissipated through each cycle because of the circuit's resistance. Our solution for the charge and the current in the circuit should look like a decaying oscillation of some kind.

To make progress, we need to do some calculating. We again assume that the capacitor initially has some charge Q_0 on it, and that we close the switch at $t = 0$. Faraday/Kirchhoff then tells us how to set up the equations for this circuit:

$$\frac{Q}{C} - IR - L\frac{dI}{dt} = 0$$

[1]Unless they are superconducting, but that opens a whole new barrel of worms ...

which reduces to

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{1}{LC}Q = 0 \ .$$

The usual way to solve these equations is the tried and true friend of the physicist, inspired guessing: we assume that the solution takes the form

$$Q(t) = e^{-t/\tau}\left(A\cos\omega t + B\sin\omega t\right)$$

— very similar to the form we found for the LC circuit. The major new feature is the exponentially decaying prefactor. We could then plug this assumed form of $Q(t)$ into the differential equation and then grind out a lot[2] of algebra. After the smoke clears, the algebra will tell you that, for this solution to work, τ and ω must take on certain values. The constants A and B will be left as adjustable parameters to be used for matching initial conditions.

Feel free to pursue this route on your own! We will instead use this moment as an opportunity to introduce a different way of thinking about this problem: we will use complex exponentials. For the remainder of this lecture, I will assume that you are familiar with the supplementary notes on complex numbers I put together (available under "Handouts" on the 8.022 webpage).

The key idea is to take the solution $Q(t)$ to be the real part of a complex function $\tilde{Q}(t)$:

$$Q(t) = \mathrm{Re}\left[\tilde{Q}(t)\right] \ .$$

Then, as long as the resistance, capacitance, and inductance are themselves real numbers, the *complex* charge $\tilde{Q}$ is also a solution to the differential equation:

$$\frac{d^2\tilde{Q}}{dt^2} + \frac{R}{L}\frac{d\tilde{Q}}{dt} + \frac{1}{LC}\tilde{Q} = 0 \ .$$

Our assumption is that

$$\tilde{Q} = Ae^{i\phi_0}e^{i\alpha t} \ .$$

As we will see shortly, A and ϕ_0 (the "amplitude" and the "phase") are the constants that we adjust to match initial conditions. We will now work out what α is. Bear in mind that α may *itself* be a complex number!!!

To proceed, we need to know the derivatives of $\tilde{Q}$. This is where the magic of using complex exponentials for solving differential equations comes in: This technique turns *differential* equations into *algebraic* ones. In particular, we have

$$\begin{aligned}\frac{d\tilde{Q}}{dt} &= i\alpha\tilde{Q} \\ \frac{d^2\tilde{Q}}{dt^2} &= -\alpha^2\tilde{Q} \ .\end{aligned}$$

Plugging this back into the governing differential equation, we get

$$\tilde{Q}\left(-\alpha^2 + i\alpha\frac{R}{L} + \frac{1}{LC}\right) = 0$$

[2]A **LOT**.

which means that α must satisfy

$$\alpha^2 - i\alpha\frac{R}{L} - \frac{1}{LC} = 0 \ .$$

This is just a quadratic equation, whose solution is

$$\alpha = i\frac{R}{2L} \pm \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \ .$$

We thus have two solutions for $\tilde{Q}$:

$$\begin{aligned}
\tilde{Q}_+ &= Ae^{i\phi_0}e^{-Rt/(2L)}e^{+it\sqrt{1/(LC)-R^2/(4L^2)}} \\
\tilde{Q}_- &= Ae^{i\phi_0}e^{-Rt/(2L)}e^{-it\sqrt{1/(LC)-R^2/(4L^2)}} \ .
\end{aligned}$$

The distinction between these two solutions is physically meaningless since they have the same real part! We might as well just keep one (the + solution, say) and be done with it. Taking the real part, our final solution becomes

$$Q(t) = Ae^{-Rt/(2L)}\cos\left(\omega t + \phi_0\right)$$

where

$$\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \ .$$

Although the use of complex numbers may seem kind of counterintuitive here, it makes solving this problem almost trivial. Rather than needing to fight through trig functions and their derivatives, you just need to solve one quadratic equation.

Complex numbers make for simple analysis.

16.3.1 The "weak damping" limit

An important limit of the above circuit is that of *weak damping*. This holds when the resistance is "small", in a well-defined way. To see what constitutes weak damping, we compute the current from our charge solution:

$$\begin{aligned}
I(t) &= -\frac{dQ}{dt} \\
&= Ae^{-Rt/(2L)}\left[\omega\sin\left(\omega t + \phi_0\right) + \frac{R}{2L}\cos\left(\omega t + \phi_0\right)\right] \ .
\end{aligned}$$

The current consists of the sum of two different sinusoidal functions. The amplitude of this first term is proportional to the frequency ω; the amplitude of the second is proportional to the damping rate $R/(2L)$.

Weak damping means that the frequency is much larger than the damping rate:

$$\omega \gg \frac{R}{2L} \ .$$

If this is the case, then the second term in the current may be ignored, to high accuracy, relative to the first term:

$$I(t) \simeq Ae^{-Rt/(2L)}\omega\sin\left(\omega t + \phi_0\right) \ .$$

We can actually make one further useful simplification in this limit: if $\omega \gg R/2L$, then you should be able to convince yourself quite easily that

$$\begin{aligned}
\omega &\equiv \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \\
&\simeq \sqrt{\frac{1}{LC}} \\
&= \omega_0 \ .
\end{aligned}$$

In other words, when we have weak damping, the oscillation frequency is — to high accuracy — the same frequency as we would have found if the resistor were not present!

In the weak damping limit, it is now simple to match the constants A and ϕ_0 to our initial conditions:

$$\begin{aligned}
Q(t=0) &= Q_0 \\
&= A\cos(\phi_0) \\
I(t=0) &= 0 \\
&= A\omega_0 \sin(\phi_0) \ .
\end{aligned}$$

The solution to these equations is

$$\begin{aligned}
A &= Q_0 \\
\phi_0 &= 0 \ .
\end{aligned}$$

Thus, our "full" solution, in the weak damping limit, for this circuit is

$$\begin{aligned}
Q(t) &= Q_0 e^{-Rt/(2L)} \cos\omega_0 t \\
I(t) &= \omega_0 Q_0 e^{-Rt/(2L)} \sin\omega_0 t \\
\omega_0 &= \sqrt{\frac{1}{LC}} \ .
\end{aligned}$$

A similar calculation can be done without assuming weak damping, but the details are much messier.

16.3.2 The quality factor

In the weak damping limit, the energy content of the circuit behaves in a very nice way. The energy of the capacitor is given by

$$\begin{aligned}
U_C(t) &= \frac{Q(t)^2}{2C} \\
&= \frac{Q_0^2}{2C} e^{-Rt/L} \cos^2\omega_0 t \ .
\end{aligned}$$

The energy of the inductor is given by

$$\begin{aligned}
U_L(t) &= \frac{1}{2} L I(t)^2 \\
&= \frac{1}{2}\omega_0^2 L Q_0^2 e^{-Rt/L} \sin^2\omega_0 t \\
&= \frac{Q_0^2}{2C} e^{-Rt/L} \sin^2\omega_0 t \ .
\end{aligned}$$

(In going from the second to the third line, I've used $\omega_0^2 = 1/LC$.) The sum of these energies is

$$\begin{aligned} U_{\rm tot}(t) &= U_C(t) + U_L(t) \\ &= \frac{Q_0^2}{2C} e^{-Rt/L}\left(\cos^2\omega_0 t + \sin^2\omega_0 t\right) \\ &= \frac{Q_0^2}{2C} e^{-Rt/L} \; . \end{aligned}$$

Since $Q_0^2/2C$ is the initial energy stored in the system, this means that the *total* energy content just exponentially damps away. (If we are not in the weak damping limit, a similar result holds, though the details are not quite as nice.)

Scientists and engineers like to be able to quantify how "good" an RLC circuit is, and so we invented a quantity called the circuit's quality Q (not to be confused with its charge). A circuit with high quality is one that oscillates many times before it loses a significant amount of energy. To make this somewhat vague notion precise, the following definition is typically adopted: "The quality factor Q is the number of radians through which a damped system oscillates by the time that its energy content has fallen by a factor of $1/e$."

This may sound a little odd, but it deconstructs into a very simple mathematical statement. First, how much time does it take for the energy to fall by a factor of $1/e$? By the above discussion, this is easy:

$$T(\text{energy to fall by } 1/e) = L/R \; .$$

Plug this time into the energy stuff discussed above, and you see that the energy is at $1/e$ of its initial level.

How many radians does the system oscillate through in this time? We just multiply the time T by the system's angular frequency $\omega \simeq \omega_0$! The quality factor of an RLC circuit is

$$\begin{aligned} Q &= \frac{\omega L}{R} \\ &\simeq \frac{\omega_0 L}{R} \; . \end{aligned}$$

As we will see when we begin exploring AC circuits, there is another way to understand the quality factor of circuits, related to how a circuit behaves "on resonance" (and very close to resonance). We will define the quality factor in what seems to be a completely different way, and will end up with exactly the same result! This is not an accident of course, but is deeply connected to how energy is stored and dissipated in circuits.

The concepts we hit in this lecture — exponential decay of an oscillation; quality factor — and that we'll hit in the next few lectures (e.g., resonance) pop up in an enormous number of contexts. For example, the quality factor of a distorted black hole — the number of radians worth of gravitational waves produced by the oscillations of its distorted event horizon emitted as it settles down to an equilibrium state — is given approximately by

$$Q_{\rm BH} \simeq 2\left(1 - \frac{c}{G}\frac{S}{M^2}\right)^{-9/20}$$

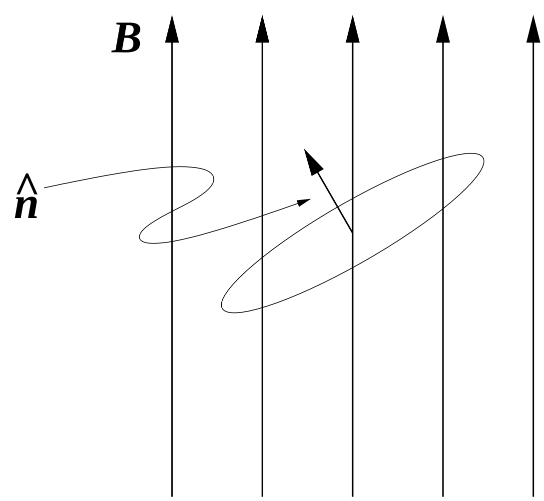
(where G is the gravitational constant, c is the speed of light, S is the magnitude of the black hole's spin angular momentum, and M is its mass). There are many other examples of quality factors that pop up repeatedly in science and engineering; you *will* see many of these notions again!

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 17:
AC Circuits; Impedance.

17.1 AC circuits: Intro

As mentioned in Lecture 15, we *use* induction to generate the vast majority of electrical power in the world. Typically, we do this with some variation on the "vary the dot product" mechanism:



The loops of wire are rotated by some external source of work — a diesel engine, a waterfall, nuclear produced steam, etc. The magnetic field largely acts as a mechanism to transform that external work into electrical power. Doing so, we end up with a driving EMF that is sinusoidal:

$$\Phi_B = BA\cos\omega t$$
$$\longrightarrow \mathcal{E} = \frac{\omega BA}{c}\sin\omega t\ .$$

At least in part for this reason, the electricity that we get "from the wall" oscillates sinusoidally. To insure some degree of uniformity, standard choices for the oscillation frequency have been agreed upon. To demonstrate that humans are illogical beings who will shoot themselves in the foot at any opportunity, two such standards are commonly used:

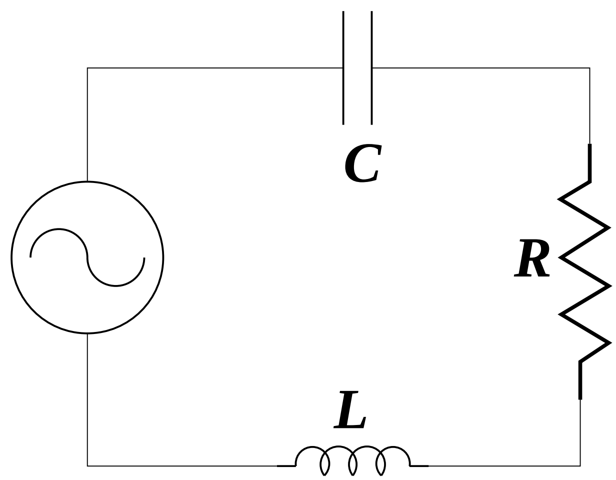
$$\begin{aligned} \omega &= 2\pi \times 60\,\mathrm{Hz} \\ \omega &= 2\pi \times 50\,\mathrm{Hz}\ . \end{aligned}$$

60 Hz is used in North America, much of South America, South Korea, and western Japan; 50 Hz is used in Europe, most of Africa (except, for example, Liberia), the rest of South America,

and most of Asia (including eastern Japan). See `http://kropla.com/electric2.htm` for a summary.

The main reason for going through this is to motivate how *essential* it is to understand how alternating current (AC) circuits work. Having spent as much time as we have studying *direct current* (DC) circuits, generalizing to AC is not too difficult — particularly if we are smart enough to recognize that our life is made a LOT easier by working with complex valued quantities.

Our goal in this lecture will be to understand how to find the current which flows in the following circuit:



The symbol on the left is a source of AC EMF. This device supplies an EMF to the rest of the circuit which we will take to be

$$\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t \ .$$

In much of our analysis, we will find it *very* useful to take this be the real part of a complex EMF $\tilde{\mathcal{E}}(t)$:

$$\mathcal{E}(t) = \mathrm{Re}\left[\tilde{\mathcal{E}}(t)\right]$$

where

$$\tilde{\mathcal{E}}(t) = \mathcal{E}_0 e^{i\omega t} \ .$$

(Note that $\mathcal{E}_0$ is a purely real number.)

We will analyze this circuit by just generalizing the operating principles that we used for DC circuits. In particular,

1. The sum of the voltage drops in the loop must equal the driving EMF. This is true for both the real form of the voltage

$$\mathcal{E}(t) = V_R(t) + V_C(t) + V_L(t)$$

and the complex form:

$$\tilde{\mathcal{E}}(t) = \tilde{V}_R(t) + \tilde{V}_C(t) + \tilde{V}_L(t) \ .$$

You may ask "What is the voltage drop across an inductor — doesn't an inductor act as a *source* of EMF?" Indeed, you would be correct for asking this. To keep things uniform, what we do is move the EMF from the left-hand side to the right-hand side, introducing a minus sign. Hence, $V_L(t) = -\mathcal{E}_{\rm ind} = L dI/dt$. There's nothing deep here — it's just a trick we introduce in order to treat all circuit elements in a uniform way.

2. The same current must pass through every circuit element. Likewise true both for real and for complex:

$$I(t) = I_R(t) = I_C(t) = I_L(t)$$
$$\tilde I(t) = \tilde I_R(t) = \tilde I_C(t) = \tilde I_L(t) \; .$$

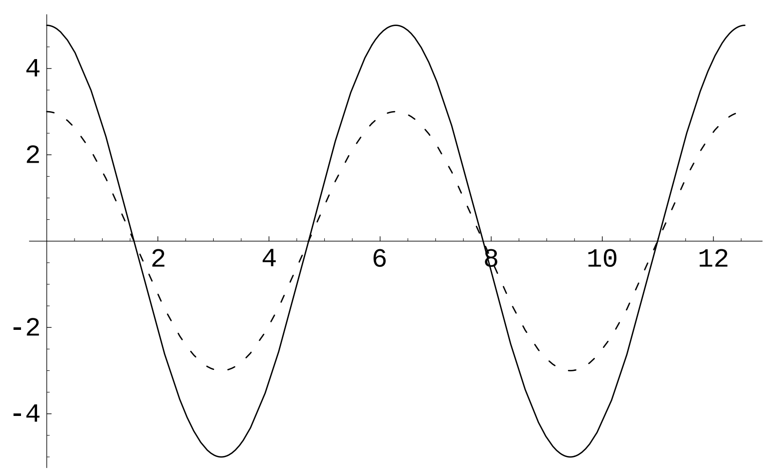
To understand this "driven RLC" circuit, we will first analyze the response of each of its circuit elements one by one.

17.1.1 AC + resistor

Resistors obey Ohm's law. Period. End of story. If we hook our driving AC EMF source across a resistor R, we have

$$\mathcal{E}(t) = V_R(t) = I(t)R \; .$$

At the risk of totally beating this to death, if we plot our driving EMF and the current in the circuit together, the plot generally looks like this:

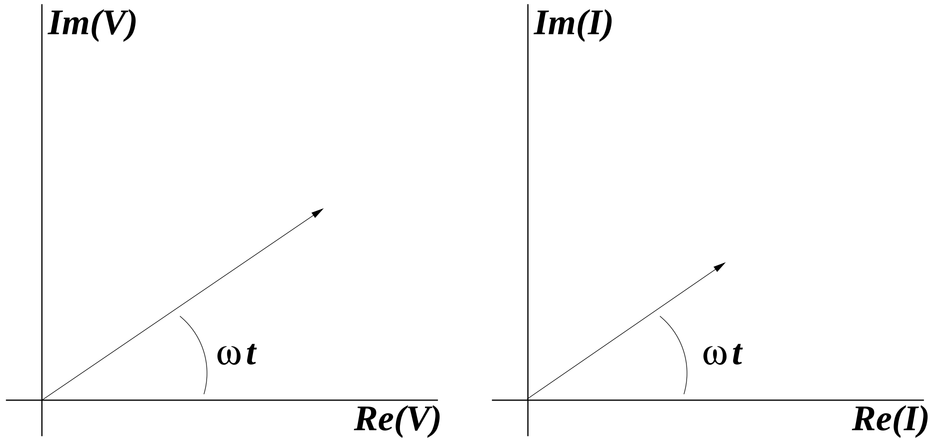


(The solid line is the driving EMF, the dotted line is the current.) The key thing to take away is the following rule:

In a resistor, the current and the voltage are IN PHASE with each other.

In other words, the peak voltage occurs at the same time as the peak current, the minimum voltage coincides with the minimum current, etc.

The same information represented by phasors in the complex plane is



"In phase" means that both phasors are at the same angle (though their magnitudes, or lengths, may be different).

17.1.2 AC + capacitor

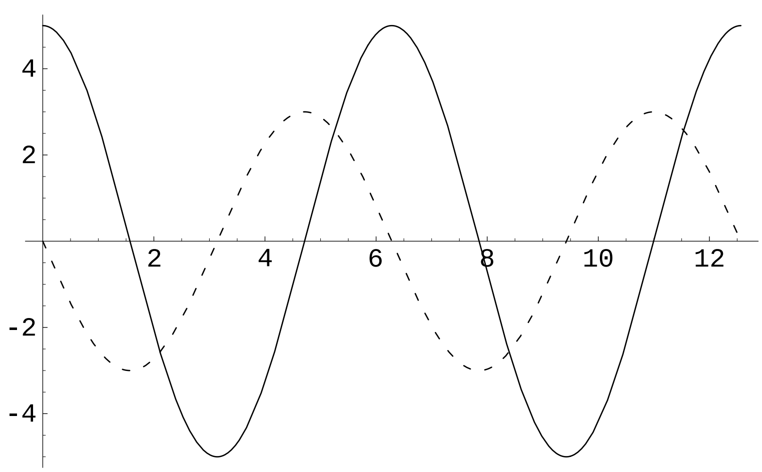
The situation is not quite so straightforward if we hook our AC EMF source across a capacitor. The governing equation in this case is

$$\mathcal{E}(t) = V_C(t) = \frac{Q(t)}{C} \; .$$

Plugging in $\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t$, solving for $Q(t)$ and taking the derivative, we have

$$\begin{aligned} I(t) &= -\omega C \mathcal{E}_0 \sin \omega t \\ &= \omega C \mathcal{E}_0 \cos(\omega t + \pi/2) \; . \end{aligned}$$

Let's plot the voltage and the current together:



Notice that maxima in the current occur before maxima in the voltage (likewise for minima):

Current LEADS voltage by 90°. Equivalently, voltage LAGS current by 90°.

The relation between the current and the voltage in the capacitor has a particularly nice form when expressed using complex numbers. The voltage is given by

$$V_C(t) = \mathcal{E}_0 \cos(\omega t) = \mathrm{Re}\left[\tilde{V}_C(t)\right]$$

where

$$\tilde{V}_C(t) = \mathcal{E}_0 e^{i\omega t} \; .$$

The current is given by

$$I(t) = \omega C \mathcal{E}_0 \cos(\omega t + \pi/2) = \mathrm{Re}\left[\tilde{I}(t)\right]$$

where

$$\begin{aligned} \tilde{I}(t) &= \omega C \mathcal{E}_0 e^{i(\omega t + \pi/2)} \\ &= i\omega C \mathcal{E}_0 e^{i\omega t} \; . \end{aligned}$$

(On the last line I have used the fact that $e^{i\pi/2} = i$.)

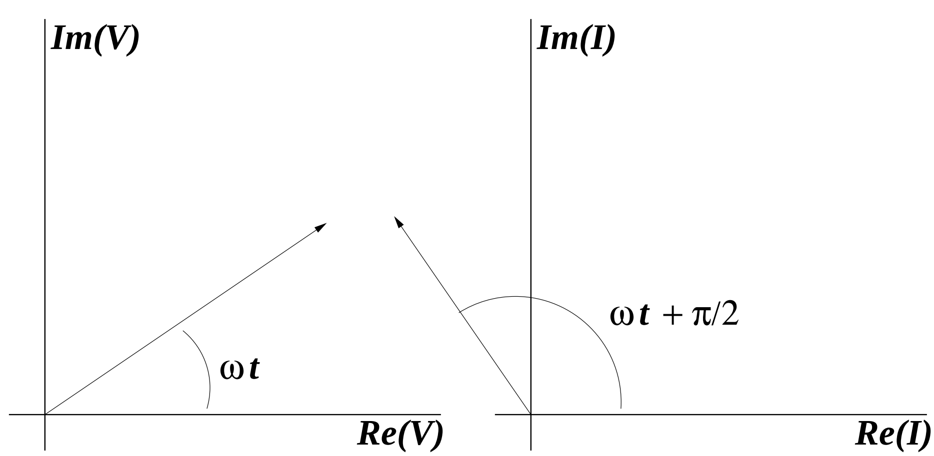
Putting all of this together, we can write the relation between the voltage and the current in the capacitor in a form that is just Ohm's law:

$$\tilde{V}_C(t) = \tilde{I}(t) Z_C$$

where

$$Z_C = \frac{1}{i\omega C}$$

is the *impedance* of the capacitor. Viewed as phasors in the complex plane, the current and the voltage look as follows:



Notice that the current is rotated 90° ahead of the voltage, in keeping with the "current leads voltage, voltage lags current" rule.

17.1.3 AC + inductor

Now, let's hook our AC EMF across an inductor. Our governing equation is clearly going to be

$$\mathcal{E}(t) = V_L(t) = L\frac{dI}{dt}\ .$$

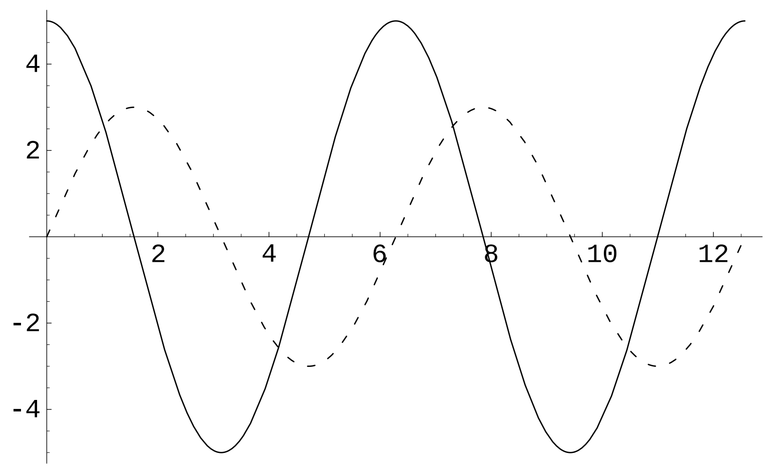
Plugging in $\mathcal{E}(t) = \mathcal{E}_0 \cos\omega t$, we find

$$\frac{dI}{dt} = \frac{\mathcal{E}_0}{L}\cos\omega t\ .$$

Integrating up, we find

$$\begin{aligned} I(t) &= \frac{\mathcal{E}_0}{\omega L}\sin\omega t \\ &= \frac{\mathcal{E}_0}{\omega L}\cos\left(\omega t - \pi/2\right)\ . \end{aligned}$$

Let's plot the voltage and the current together:



In this case, maxima in the current occur *after* maxima in the voltage (likewise for minima):

Current LAGS voltage by $90°$. Equivalently, voltage LEADS current by $90°$.

The relation between the inductor's current and voltage again has a particularly nice form when expressed with complex numbers. The voltage is

$$V_L(t) = \mathcal{E}_0\cos(\omega t) = \mathrm{Re}\left[\tilde{V}_L(t)\right]$$

where

$$\tilde{V}_L(t) = \mathcal{E}_0 e^{i\omega t}\ .$$

The current is given by

$$I(t) = \frac{\mathcal{E}_0}{\omega L}\cos(\omega t - \pi/2) = \mathrm{Re}\left[\tilde{I}(t)\right]$$

where

$$\begin{aligned}\tilde{I}(t) &= \frac{\mathcal{E}_0}{\omega L} e^{i(\omega t - \pi/2)} \\ &= \frac{\mathcal{E}_0}{i\omega L} e^{i\omega t} \,.\end{aligned}$$

(On the last line I have used the fact that $e^{-i\pi/2} = 1/i$.)

Putting all of this together, we can write the relation between the voltage and the current in the inductor in a form that is just Ohm's law:

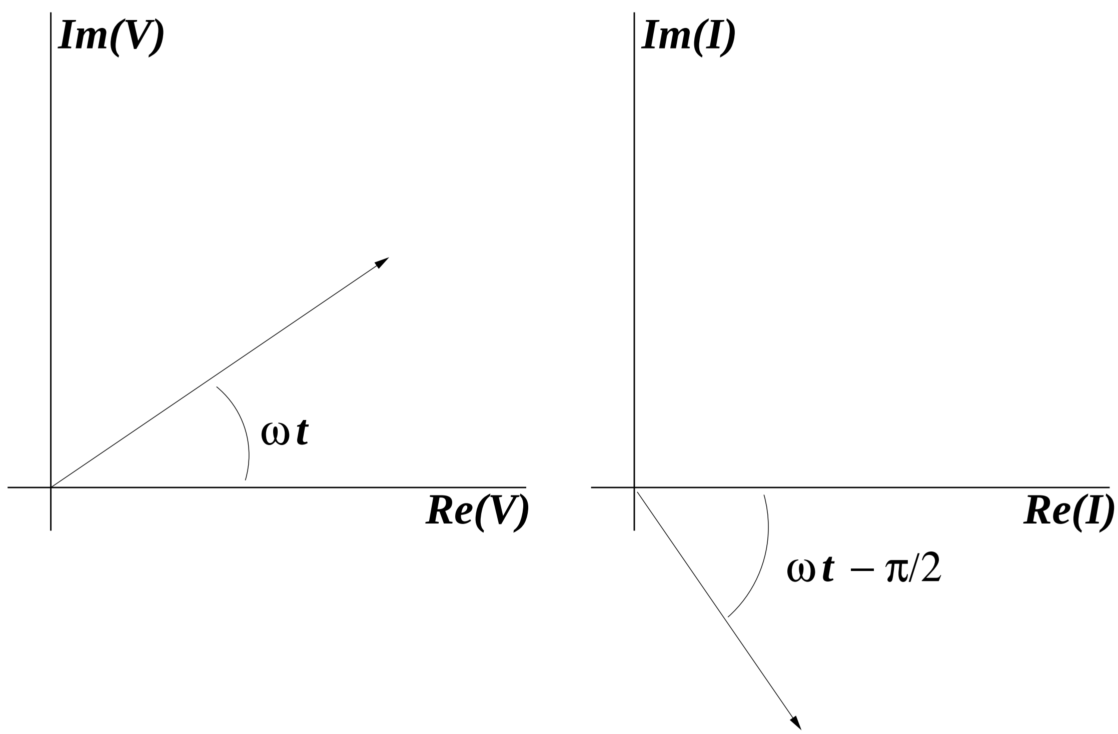
$$\tilde{V}_L(t) = \tilde{I}(t) Z_L$$

where

$$Z_L = i\omega L$$

is the impedance of the inductor.

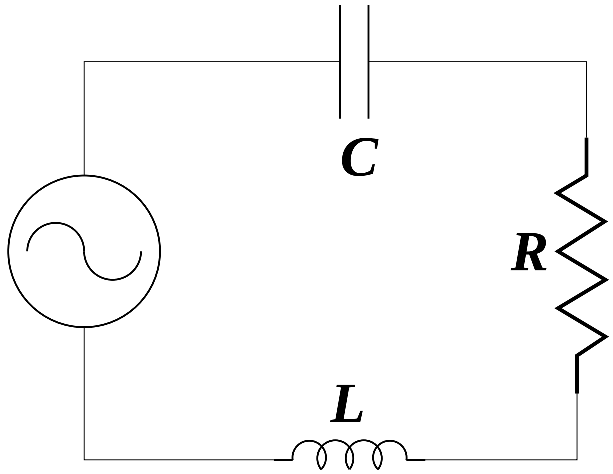
Viewed as phasors in the complex plane, the current and the voltage are



Notice that the current is 90° behind in the voltage, in keeping with the notion that current lags voltage in inductors.

17.2 Putting it together: Driven series RLC, take 1

We're now finally ready to analyze this circuit:



Doing so is actually quite easy as long as we work in the complex domain! We just use the rules we established earlier,

$$\tilde{\mathcal{E}}(t) = \tilde{V}_R(t) + \tilde{V}_L(t) + \tilde{V}_C(t)$$
$$\tilde{I}(t) = \tilde{I}_R(t) = \tilde{I}_L(t) = \tilde{I}_C(t)$$

to which we now add our new version of Ohm's law:

$$\tilde{V}_X(t) = \tilde{I}_X(t) Z_X$$

where X stands for the resistor, the capacitor, or the inductor.

Our first equation becomes

$$\begin{aligned} \mathcal{E}_0 e^{i\omega t} &= \tilde{I}(t)\left[R + i\omega L + \frac{1}{i\omega C}\right] \\ &= \tilde{I}(t)\left[R + i\left(\omega L - \frac{1}{\omega C}\right)\right] \\ &= \tilde{I}(t) Z_{\text{tot}} \end{aligned}$$

where Z_{tot} is the *total* impedance of the circuit:

$$Z_{\text{tot}} = R + i\omega L - \frac{i}{\omega C} = R + i\left(\omega L - \frac{1}{\omega C}\right) .$$

Notice that the total impedance is dominated by the inductor at high frequency, and by the capacitor at low frequency. We will see this again very soon.

Using the above result, the *complex* current in the circuit is given by

$$\tilde{I}(t) = \frac{\mathcal{E}_0 e^{i\omega t}}{R + i\left[\omega L - 1/(\omega C)\right]} .$$

To keep our notation simple, we will write this as

$$\tilde{I}(t) = I_0 e^{i\omega t} e^{-i\phi} .$$

In other words, we are just factoring out the $e^{i\omega t}$ piece, and putting

$$I_0 e^{-i\phi} = \frac{\mathcal{E}_0}{R + i\left[\omega L - 1/(\omega C)\right]}\ .$$

Don't get excited about the minus sign in front of the $i\phi$. All this means is that we *define* ϕ as the angle by which the current will **LAG** the voltage. This is convention, not anything deep. If ϕ turns out to be negative, then it means that the current actually leads the voltage.

Our job now will be to figure out the "current amplitude" I_0 and the current phase angle ϕ. To make progress on this, we note that *any* complex number can be written in the form

$$z = re^{-i\phi}$$

where $r = \sqrt{zz^*}$. Let's look at the magnitude of the current first:

$$\begin{aligned}
|\tilde{I}| = I_0 &= \sqrt{\tilde{I}\tilde{I}^*} \\
&= \mathcal{E}_0 \sqrt{\left(\frac{1}{R + i\left[\omega L - 1/(\omega C)\right]}\right)\left(\frac{1}{R - i\left[\omega L - 1/(\omega C)\right]}\right)} \\
&= \frac{\mathcal{E}_0}{\sqrt{R^2 + \left[\omega L - 1/(\omega C)\right]^2}}\ .
\end{aligned}$$

This is pretty cool. The amount of current that flows through this circuit varies — quite strongly! — as a function of the AC frequency ω. Notice that the *maximum* current amplitude occurs when $\omega = \omega_0 = 1/\sqrt{LC}$: at this frequency, the L and the C terms cancel each other out, leaving just the resistance term.

Now we need to figure out that phase term. To set things up, note that from the rule

$$z = re^{-i\phi} = r\left(\cos\phi - i\sin\phi\right)$$

we can extract

$$\tan\phi = -\frac{\mathrm{Im}\left[z\right]}{\mathrm{Re}\left[z\right]}\ .$$

Now, we want to do something like this for our complex current. Let us massage this relation a bit:

$$\begin{aligned}
I_0 e^{-i\phi} &= \frac{\mathcal{E}_0}{R + i\left[\omega L - 1/(\omega C)\right]} \\
&= \frac{\mathcal{E}_0}{R + i\left[\omega L - 1/(\omega C)\right]}\left(\frac{R - i\left[\omega L - 1/(\omega C)\right]}{R - i\left[\omega L - 1/(\omega C)\right]}\right) \\
&= \frac{\mathcal{E}_0}{R^2 + \left[\omega L - 1/(\omega C)\right]^2}\left(R - i\left[\omega L - 1/(\omega C)\right]\right)\ .
\end{aligned}$$

Taking the ratio of the imaginary part to the real part, we have

$$\begin{aligned} \tan\phi &= \frac{\omega L - 1/(\omega C)}{R} \\ &= \frac{\omega L}{R} - \frac{1}{\omega C R} \,. \end{aligned}$$

Typically, ϕ ends up having a shape as a function of ω that looks something like this:

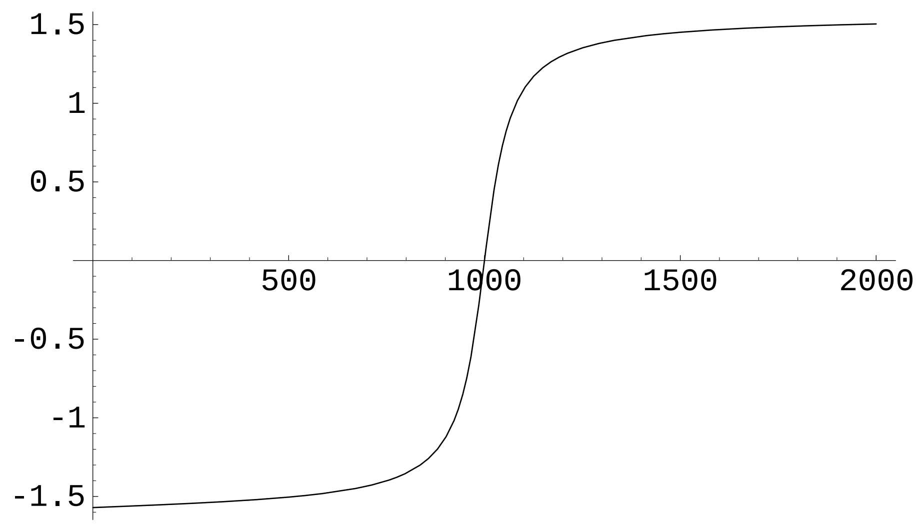
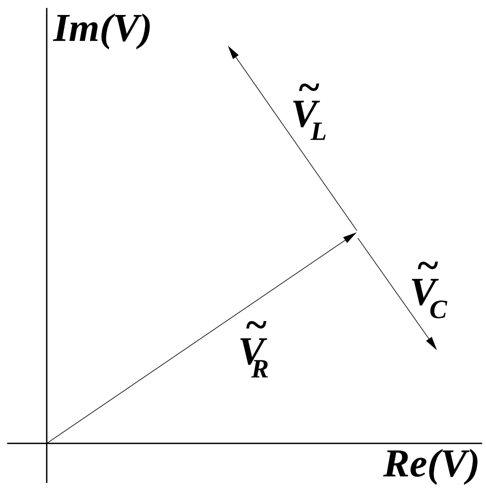


Figure 1: ϕ vs ω for $R = 1$ Ohm, $L = 0.01$ Henry, $C = 10^{-4}$ Farad.

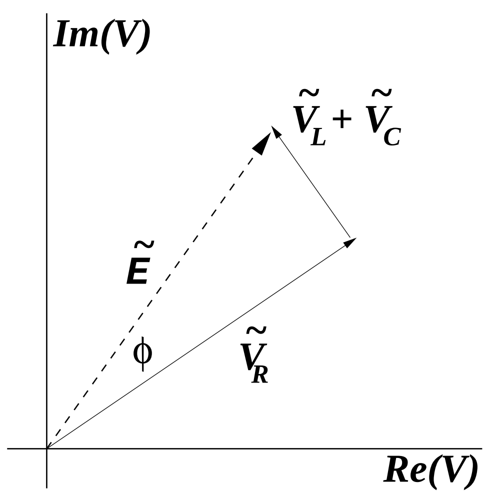
Notice that, for low frequency, the phase angle limits to $-\pi/2$: the current *LEADS* the voltage by $\sim 90°$ at low frequency. This is exactly the behavior we saw when we considered our EMF attached to just the capacitor! This has a very nice interpretation: at low frequencies, the capacitor dominates the overall impedance of the circuit — exactly what we expected on intuitive grounds from the form of $Z_{\rm tot}$. Likewise, at high frequencies, the phase angle limits to $+\pi/2$: the current *LAGS* voltage by $\sim 90°$ — exactly the behavior we see for EMF attached to the inductor; and, exactly what $Z_{\rm tot}$ would suggest.

17.3 Putting it together: Driven series RLC, take 2

The phasor picture allows us to put together another way of getting the above result that is perhaps a little more intuitive. The basic idea is to treat the phasors as vectors in the complex plane. Then, we just add them all head to tail:



This diagram shows the voltage drop across the resistor, $\tilde{V}_R$, along with the drop across the inductor (90° ahead) and across the capacitor (90° behind). This drawing simplifies by combining the voltage drops across the inductor and the capacitor:



We can now work out a lot of things using simple geometry and trigonometry! For example, the rule that the total voltage drop in the circuit must equal the driving EMF tells us that

$$\tilde{V}_R + \tilde{V}_L + \tilde{V}_C = \tilde{\mathcal{E}} \ .$$

The driving EMF $\tilde{\mathcal{E}}$ is the "resultant" vector we obtain by summing the other three phasors; ϕ is the angle between $\tilde{\mathcal{E}}$ and $\tilde{V}_R$ (since $\tilde{V}_R$ is in phase with the current in the circuit).

With the Pythagorean theorem, we get the current's magnitude very easily:

$$\mathcal{E}_0^2 = V_R^2 + (V_L - V_C)^2$$

$$
\begin{aligned}
&= I_0^2\left[R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2\right] \\
\longrightarrow I_0 &= \frac{\mathcal{E}_0}{\sqrt{R^2 + [\omega L - 1/(\omega C)]^2}} \,.
\end{aligned}
$$

Likewise, we can get the phase angle ϕ with simple trigonometry:

$$
\begin{aligned}
\tan\phi &= \frac{V_L - V_C}{V_R} \\
&= \frac{\omega L}{R} - \frac{1}{\omega C R} \,.
\end{aligned}
$$

Exactly what we had before!

17.4 Summary

AC circuits are surprisingly easy to analyze, provided you follow some simple rules:

1. Do all of your calculations using complex voltages and currents. Replace $\cos(\omega t)$ with $e^{i\omega t}$; replace $\sin(\omega t)$ with $\cos(\omega t - \pi/2) = e^{i(\omega t - \pi/2)}$.

2. Relate the voltage drops and currents across any circuit element using our "generalized Ohm's law":

$$\tilde{V}_X = \tilde{I}_X Z_X \,.$$

The quantity Z_X is the impedance of a circuit element X. Impedence is just resistance on steroids: it works exactly like resistance in a DC circuit, but is complex, frequency dependent, and works in AC circuits. The impedance formulas you need to know are

$$
\begin{aligned}
Z_R &= R \\
Z_C &= 1/(i\omega C) \\
Z_L &= i\omega L \,.
\end{aligned}
$$

3. Analyze your circuit using impedance and complex voltages and currents just like you would analyze a DC circuit containing only resistors. Be careful to get all magnitudes and phases correct — the math is likely to be a little more hairy than a simple DC circuit. A good phasor diagram can help organize things.

4. Take the real parts of all relevant quantities at the end of the day.

And that's it.

Scott Hughes *21 April 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 18:
AC CIRCUITS.
POWER AND ENERGY; RESONANCE. FILTERS.

18.1 AC circuits: Recap and summary

Last time, we looked at AC circuits and found that they are quite simple to analyze provided we follows some simple rules:

1. Work with complex valued voltages and currents. Our driving AC EMF is usually something like $\mathcal{E}(t) = \mathcal{E}_0 \cos\omega t$; replace this with $\tilde{\mathcal{E}}(t) = \mathcal{E}_0 e^{i\omega t}$ (so that $\mathcal{E}(t) = \mathrm{Re}[\tilde{\mathcal{E}}_0(t)]$).

2. The voltage drop across any circuit element obeys a generalized, complex version of Ohm's law:

$$\tilde{V}_X = \tilde{I}_X Z_X$$

where Z_X is the impedance of circuit element X. Impedance works just like resistance, but is complex and frequency dependent:

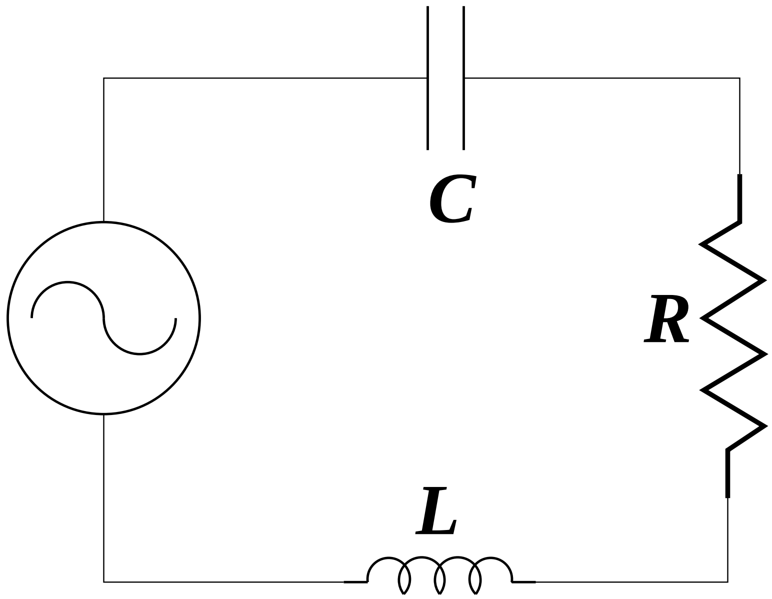
$$\begin{aligned} Z_R &= R \\ Z_C &= 1/(i\omega C) \\ Z_L &= i\omega L\ . \end{aligned}$$

3. Analyze the circuit as though it were a simple DC circuit containing only resistors as circuit elements. You may find phasor diagrams helpful for making sure that you get the magnitudes and phases correct.

4. Take the real part at the end of the day.

18.2 Power delivered to an AC circuit

Let's look again at our prototypical driven RLC circuit:



Assume that the EMF supplied is $\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t$. In the complex representation, this becomes $\tilde{\mathcal{E}}(t) = \mathcal{E}_0 e^{i\omega t}$.

We can now solve for the complex current in this circuit:

$$\begin{aligned} \tilde{\mathcal{E}} &= \tilde{V}_R + \tilde{V}_L + \tilde{V}_C \\ &= \tilde{I}\left[R + i\omega L + \frac{1}{i\omega C}\right] \\ \longrightarrow \tilde{I} &= \frac{\tilde{\mathcal{E}}}{R + i[\omega L - 1/(\omega C)]} \ . \end{aligned}$$

We now substitute $\tilde{\mathcal{E}}(t) = \mathcal{E}_0 e^{i\omega t}$, $\tilde{I}(t) = I_0 e^{-i\phi} e^{i\omega t}$. Then, we find

$$\begin{aligned} I_0 &= \frac{\mathcal{E}_0}{\sqrt{R^2 + [\omega L - 1/(\omega C)]^2}} \\ &\equiv \mathcal{E}_0/|Z_{\rm tot}(\omega)| \end{aligned}$$

where $|Z_{\rm tot}(\omega)|$ is the magnitude of the total impedance. The phase is defined by

$$\begin{aligned} \tan\phi &= -\frac{\mathrm{Im}[\tilde{I}]}{\mathrm{Re}[\tilde{I}]} \\ &= \frac{\omega L}{R} - \frac{1}{\omega C R} \ . \end{aligned}$$

The *real* current flowing in this circuit is thus given by

$$I(t) = I_0 \cos(\omega t - \phi)$$

where

$$\begin{aligned}
I_0 &= \frac{\mathcal{E}_0}{|Z_{\rm tot}(\omega)|} \\
\tan\phi &= \frac{\omega L}{R} - \frac{1}{\omega C R}\ .
\end{aligned}$$

How much power is delivered to this circuit? From the definition of EMF as work per unit charge, and from the definition of current as charge per unit time, we know that the power delivered to this circuit must be

$$P(t) = I(t)\mathcal{E}(t)\ .$$

Let's plug in the results we have for $I(t)$ and $\mathcal{E}(t)$:

$$\begin{aligned}
P(t) &= \frac{\mathcal{E}_0^2}{|Z_{\rm tot}(\omega)|}\cos\omega t\cos(\omega t - \phi) \\
&= \frac{\mathcal{E}_0^2}{|Z_{\rm tot}(\omega)|}\left[\cos\omega t\cos\omega t\cos\phi + \cos\omega t\sin\omega t\sin\phi\right]\ .
\end{aligned}$$

On the second line, I have used a trig identity, $\cos(a-b) = \cos a\cos b + \sin a\sin b$, to expand the second cosine.

This result for $P(t)$ is exact. It tells us that the power delivered to the circuit is a quantity that oscillates. In many circumstances, the frequency of this oscillation is *far* too rapid for us to be able to track the detailed oscillations. For example, electricity out of the wall oscillates 60 times per second — way faster than anything we can keep track of without special equipment.

In this circumstance, when the oscillations are much faster than the phenomena that we are interested in, it makes a lot more sense to *average*. The "average" of any oscillating function is given by integrating that function over its period, and then dividing by the period. For example, the average power delivered to the circuit is given by

$$\begin{aligned}
\langle P\rangle &= \frac{1}{T}\int_0^T P(t)\,dt \\
&= \frac{\omega}{2\pi}\int_0^{2\pi/\omega} P(t)\,dt\ .
\end{aligned}$$

(On the last line, I'm using the fact that the period of a function whose angular frequency is ω is $2\pi/\omega$.)

The average power that the EMF delivers to the circuit is thus given by

$$\langle P\rangle = \frac{\mathcal{E}_0^2}{|Z_{\rm tot}(\omega)|}\frac{\omega}{2\pi}\left[\int_0^{2\pi/\omega}\left(\cos\omega t\cos\omega t\cos\phi\right)\,dt + \int_0^{2\pi/\omega}\left(\cos\omega t\sin\omega t\sin\phi\right)\,dt\right]\ .$$

The integrals we need are

$$\begin{aligned}
\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\left(\cos\omega t\cos\omega t\cos\phi\right)\,dt &= \frac{1}{2}\cos\phi \\
\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\left(\cos\omega t\sin\omega t\cos\phi\right)\,dt &= 0\ .
\end{aligned}$$

The average power delivered to the circuit is thus

$$\langle P(\omega)\rangle = \frac{1}{2}\frac{\mathcal{E}_0^2\cos\phi}{|Z_{\rm tot}(\omega)|}\ .$$

(I've written the power as a function of ω to emphasize its dependence on the driving frequency.) This average power is the one that is used to describe most common household items. For example, when we talk about a 100 Watt lightbulb, we mean that the average power delivered to the bulb is 100 Watts.

This power is often written in terms of the "root mean squared", or RMS, voltage. The RMS is often used when we talk about oscillating quantities whose average values are zero: we average the *square* of the quantity. For example, the RMS EMF is given by

$$\begin{aligned}
\mathcal{E}_{\rm rms}^2 &= \frac{\omega}{2\pi}\int_0^{2\pi/\omega}\mathcal{E}(t)^2\,dt \\
&= \frac{\mathcal{E}_0^2}{2} \\
\longrightarrow \mathcal{E}_{\rm rms} &= \frac{\mathcal{E}_0}{\sqrt{2}}\ .
\end{aligned}$$

For purely sinusoidal AC EMF and current, the RMS value is just the amplitude divided by $\sqrt{2}$. Typically, when AC voltages are discussed, it is the RMS value that is indicated. For example, the wall voltage throughout North America is typically 120 Volts. This is the RMS value of the voltage from the wall; the *amplitude* of the voltage is $\sqrt{2}\times 120 = 170$ Volts. Note that, since $I_0 = \mathcal{E}_0/|Z_{\rm tot}|$, we also have $I_{\rm rms} = \mathcal{E}_{\rm rms}/|Z_{\rm tot}|$.

The average power delivered by the AC EMF can thus be written

$$\langle P(\omega)\rangle = \frac{\mathcal{E}_{\rm rms}^2\cos\phi}{|Z_{\rm tot}(\omega)|}\ .$$

This is the form in which we most commonly see the AC power related to AC voltage (or currents).

With a little bit of effort, we can show that

$$\cos\phi = \frac{R}{\sqrt{R^2 + [\omega L - 1/(\omega C)]^2}} = \frac{R}{|Z_{\rm tot}(\omega)|}\ .$$

Using this, the average power can be rewritten as follows:

$$\begin{aligned}
\langle P(\omega)\rangle &= \frac{\mathcal{E}_{\rm rms}^2 R}{|Z_{\rm tot}(\omega)|^2} \\
&= I_{\rm rms}(\omega)^2\,R\ .
\end{aligned}$$

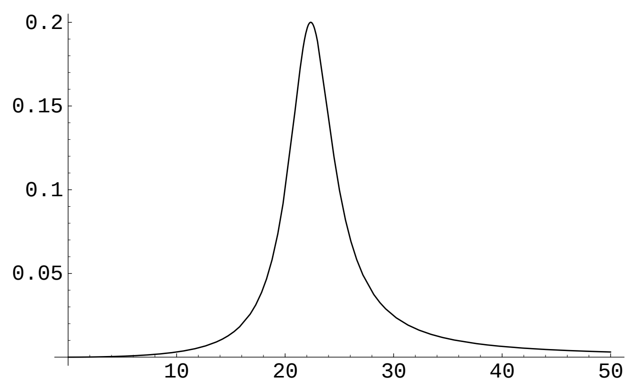
The average power delivered has a *really* simple form when expressed in terms of the RMS current — it looks just like the DC power dissipation formula! (Be a little careful, though: don't forget that the current in this case, $I_{\rm rms}$, is a function of ω.)

18.3 Resonance

The average power that is delivered to the circuit is a function of frequency. Using the formula

$$\langle P(\omega)\rangle = \frac{\mathcal{E}_{\rm rms}^2 R}{|Z_{\rm tot}(\omega)|^2}$$

we see that the frequency dependence enters through $1/|Z_{\rm tot}|^2$. This has the following appearance as a function of ω:



(The parameters $R = 5$ Ohms, $L = 1$ Henry, $C = 2 \times 10^{-3}$ Farads were used for this plot.) Clearly, there is some frequency at which the amount of power delivered is a maximum — at this frequency, the circuit is *in resonance* with the driving EMF. Our goal now will be to calculate this frequency: we look for maxima of

$$\frac{1}{|Z_{\rm tot}(\omega)|^2} = \frac{1}{R^2 + [\omega L - 1/(\omega C)]^2}\ .$$

We could do the usual procedure of looking for frequencies at which the derivative of this with respect to ω goes to zero; however, the answer is actually obvious by inspection! The maximum occurs when the ωL and $1/\omega C$ terms cancel each other out. This happens when

$$\omega = \omega_{\rm res} = \frac{1}{\sqrt{LC}}\ .$$

Notice that the resonant frequency is ω_0, the oscillation frequency of the pure LC circuit! At this frequency, $|Z_{\rm tot}| = R$, and so

$$\langle P(\omega_{\rm res})\rangle = \langle P_{\rm max}\rangle = \frac{\mathcal{E}_{\rm rms}^2}{R}\ .$$

Also at this frequency, we must have $\phi = 0$ — the current in the circuit is exactly in phase with the driving EMF. The voltage drops due to the capacitor and the inductor precisely cancel each other out — it is as though only the resistor matters in the circuit.

When one talks about a resonant system, one would like to quantify how "good" the resonance is. A "good" resonance is one in which the power versus ω curve is very narrow — the power delivered has a very sharp peak. A convenenient way to quantify this is to look for the two frequencies, $\omega_{\rm high}$ and $\omega_{\rm low}$, at which the power delivered has fallen to 1/2 of the peak value:

$$\langle P(\omega_{\rm high})\rangle = \langle P(\omega_{\rm low})\rangle = \frac{1}{2}\langle P_{\rm max}\rangle\ .$$

This tells us that

$$\begin{aligned}\frac{1}{2R} &= \frac{R}{|Z_{\rm tot}(\omega_{\rm high/low})|^2}\\ &= \frac{R}{R^2 + [\omega_{\rm high/low}L - 1/(\omega_{\rm high/low}C)]^2}\ .\end{aligned}$$

This in turn tells us that

$$\omega_{\rm high/low}L - \frac{1}{\omega_{\rm high/low}C} = R\ .$$

This equation can be rearranged to form a quadratic equation; solving it yields the two frequencies $\omega_{\rm high}$ and $\omega_{\rm low}$.

The "width" of the peak $\Delta\omega$ is then defined as $\Delta\omega = \omega_{\rm high} - \omega_{\rm low}$. The ratio of this width to the resonant frequency is then defined as the *quality factor* of the driven system:

$$Q \equiv \frac{\omega_{\rm res}}{\Delta\omega}\ .$$

On an upcoming problem set, you will show that limit,

$$\Delta\omega = \frac{R}{L}\ .$$

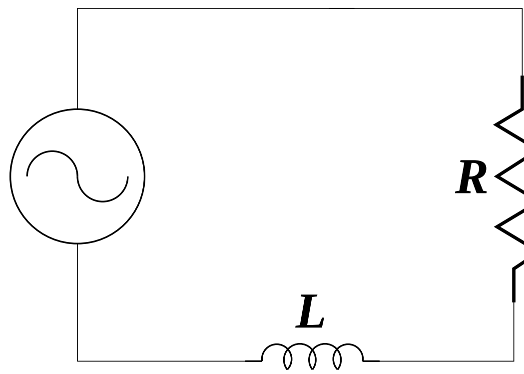
Combining this with the fact that $\omega_{\rm res} = \omega_0$ (the oscillation frequency of the LC circuit) and we have

$$Q = \frac{\omega_0 L}{R}\ .$$

This is *exactly* the same quality factor that we determined for the *un*-driven RLC circuit! This is not an accident. In the undriven system, the quality factor tells us about how quickly the system loses energy. In the driven system, it tells us about the power delivered to the system on resonance compared to slightly off resonance. Although it is beyond the scope of this course to prove this, these two concepts are intimately related to one another.

18.4 Filters

The frequency dependent response of an RLC circuit means that they can act as *filters* — different frequencies are preferentially "let through" different parts of the circuit. To see this principle in action, consider the following simple LR circuit with an AC driving EMF:



Working in the complex representation, we know that

$$\tilde{\mathcal{E}} = \tilde{V}_R + \tilde{V}_L \ .$$

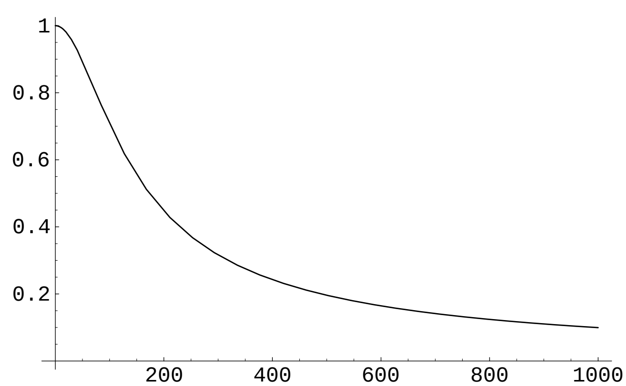
We also know there is a single current passing through both the resistor and the inductor:

$$\tilde{\mathcal{E}} = \tilde{I}\,(R + i\omega L)$$
$$\tilde{I} = \frac{\tilde{\mathcal{E}}}{R + i\omega L} \ .$$

What then is the *magnitude* of the voltage drop across the resistor? Multiplying by R and taking the magnitudes, we find

$$V_R = |\tilde{I}R| = \mathcal{E}\frac{R}{\sqrt{R^2 + \omega^2 L^2}} \ .$$

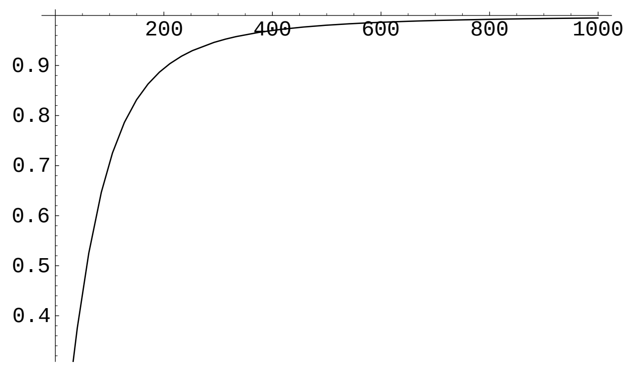
In this circuit, the resistor acts like a *low pass filter*: the voltage at low frequencies is essentially the same magnitude as the "input" voltage. However, at high frequencies, the voltage is very suppressed. Plotting V_R versus ω, we find something like this:



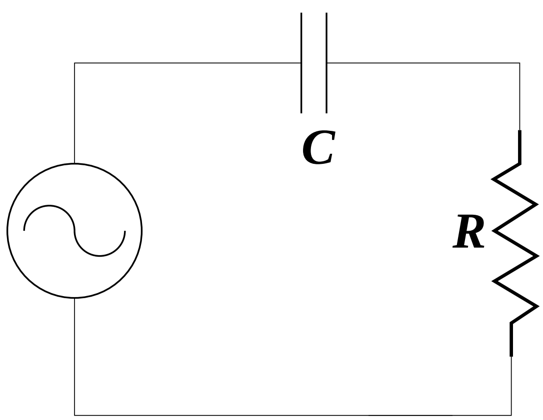
How about the magnitude of the voltage drop across the inductor? In this case, we have

$$V_L = |\tilde{I} Z_L| = \mathcal{E}\frac{\omega L}{\sqrt{R^2 + \omega^2 L^2}}\ .$$

The inductor acts like a *high pass filter*: high frequency parts of the signal get through, but the stuff at low frequencies is highly suppressed: Plotting V_L versus ω gives us



How about if we have a circuit like this:



This again gives us filtering behavior. The complex current in the circuit is given by

$$\begin{aligned}\tilde{I} &= \frac{\tilde{\mathcal{E}}}{R + 1/(i\omega C)}\\ &= \frac{i\omega C\tilde{\mathcal{E}}}{1 + i\omega RC}\ .\end{aligned}$$

The voltage across the resistor is

$$V_R = |\tilde{I} R| = \mathcal{E}\frac{\omega CR}{\sqrt{1 + \omega^2 R^2 C^2}}\ .$$

This acts like a *high* pass filter: $V_R = 0$ at $\omega = 0$, but asymptotically approaches $\mathcal{E}$ as we go to high frequency.

The voltage across the capacitor is

$$V_C = |\tilde{I} Z_C| = \mathcal{E} \frac{1}{\sqrt{1 + \omega^2 R^2 C^2}} \, .$$

This acts as a *low* pass filter.

The rule that capacitors act as low pass devices and inductors as high pass devices follows quite simply from the form of their impedance: A capacitor has a large impedance at low frequency,

$$Z_C = 1/(i\omega C)$$

and so shows a large voltage drop at low frequency. Likewise, an inductor has a large impedance at high frequency,

$$Z_L = i\omega L$$

and hence shows a large voltage drop at high frequency.

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 19:
Displacement current. Maxwell's equations.

19.1 Inconsistent equations

Over the course of this semester, we have derived 4 relationships between the electric and magnetic fields on the one hand, and charge and current density on the other. They are:

$$
\begin{aligned}
\text{Gauss's law:} &\qquad \vec{\nabla}\cdot\vec{E} = 4\pi\rho \\
\text{Magnetic law:} &\qquad \vec{\nabla}\cdot\vec{B} = 0 \\
\text{Faraday's law:} &\qquad \vec{\nabla}\times\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t} \\
\text{Ampere's law:} &\qquad \vec{\nabla}\times\vec{B} = \frac{4\pi\vec{J}}{c}
\end{aligned}
$$

As written, these equations are slightly inconsistent. We can see this inconsistency very easily — we just take the divergence of both sides of Ampere's law. Look at the left hand side first:

$$\text{Left hand side:}\qquad \vec{\nabla}\cdot\left(\vec{\nabla}\times\vec{B}\right) = 0\ .$$

This follows from the rule that the divergence of the curl is always zero (as you proved on pset 3). Now look at the right hand side:

$$\text{Right hand side:}\qquad \vec{\nabla}\cdot\left(\frac{4\pi\vec{J}}{c}\right) = -\frac{4\pi}{c}\frac{\partial\rho}{\partial t}\ .$$

Here, I've used the continuity equation, $\vec{\nabla}\cdot\vec{J} = -\partial\rho/\partial t$.

Ampere's law is inconsistent with the continuity equation except when $\partial\rho/\partial t = 0$!!! A charge density that is constant in time is actually a fairly common circumstance in many applications, so it's not too surprising that we can go pretty far with this "incomplete" version of Ampere's law. But, as a matter of principle — and, as we shall soon see, of practice as well — it's just not right. We need to fix it somehow.

19.2 Fixing the inconsistency

We can fix up this annoying little inconsistency by inspired guesswork. When we take the divergence of Ampere's left hand side, we get zero — no uncertainty about this whatsoever, it's just ZERO. We should be able to add a function to the right hand side such that the divergence of the right side is forced to be zero as well.

Let's suppose that our "generalized Ampere's law" takes the form

$$\vec{\nabla}\times\vec{B} = \frac{4\pi\vec{J}}{c} + \vec{\mathcal{F}}\ .$$

Our goal now is to figure out what $\vec{\mathcal{F}}$ must be. Taking the divergence of both sides, we find

$$\begin{aligned} 0 &= \frac{4\pi\vec{\nabla}\cdot\vec{J}}{c} + \vec{\nabla}\cdot\vec{\mathcal{F}} \\ \longrightarrow \vec{\nabla}\cdot\left(c\vec{\mathcal{F}}\right) &= 4\pi\frac{\partial\rho}{\partial t} \, . \end{aligned}$$

Our mystery function has the property that when we take its divergence (and multiply by c), we get the rate of change of charge density.

This looks a lot like Gauss's law! If we take the time derivative of Gauss's law, we have

$$\begin{aligned} \frac{\partial}{\partial t}\vec{\nabla}\cdot\vec{E} &= 4\pi\frac{\partial\rho}{\partial t} \\ \longrightarrow 4\pi\frac{\partial\rho}{\partial t} &= \vec{\nabla}\cdot\left(\frac{\partial\vec{E}}{\partial t}\right) \end{aligned}$$

On the last line, I've use the fact that it is OK to exchange the order of partial derivatives:

$$\begin{aligned} \frac{\partial}{\partial t}\vec{\nabla} &= \hat{x}\frac{\partial}{\partial t}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial t}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial t}\frac{\partial}{\partial z} \\ &= \hat{x}\frac{\partial}{\partial x}\frac{\partial}{\partial t} + \hat{y}\frac{\partial}{\partial y}\frac{\partial}{\partial t} + \hat{z}\frac{\partial}{\partial z}\frac{\partial}{\partial t} \\ &= \vec{\nabla}\frac{\partial}{\partial t} \, . \end{aligned}$$

(This actually works in any coordinate system; Cartesian coordinates are good enough to demonstrate the point.)

Substituting in for $4\pi\,\partial\rho/\partial t$, we have

$$\vec{\nabla}\cdot\left(c\vec{\mathcal{F}}\right) = \vec{\nabla}\cdot\left(\frac{\partial\vec{E}}{\partial t}\right)$$

which tells us

$$\vec{\mathcal{F}} = \frac{1}{c}\frac{\partial\vec{E}}{\partial t} \, .$$

Ampere's law becomes

$$\vec{\nabla}\times\vec{B} = \frac{4\pi\vec{J}}{c} + \frac{1}{c}\frac{\partial\vec{E}}{\partial t} \, .$$

Substituting this in with our other equations yields the *Maxwell equations*:

$$\begin{aligned} \text{Gauss's law:} \qquad & \vec{\nabla}\cdot\vec{E} = 4\pi\rho \\ \text{Magnetic law:} \qquad & \vec{\nabla}\cdot\vec{B} = 0 \\ \text{Faraday's law:} \qquad & \vec{\nabla}\times\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t} \\ \text{Generalized Ampere's law:} \qquad & \vec{\nabla}\times\vec{B} = \frac{4\pi\vec{J}}{c} + \frac{1}{c}\frac{\partial\vec{E}}{\partial t} \, . \end{aligned}$$

It was James Clerk Maxwell who first fixed this inconsistency and argued that the $\partial \vec{E}/\partial t$ term must be present in Ampere's law. Interestingly, he did not argue this on the basis of continuity, but rather purely on the grounds of symmetry. We'll return to this point later.

You should also be aware of the form of these equations in SI units[1] — all those cs and 4πs are replaced with various combinations of μ_0s and ϵ_0s:

$$
\begin{aligned}
\text{Gauss's law:} \qquad & \vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0} \\
\text{Magnetic law:} \qquad & \vec{\nabla} \cdot \vec{B} = 0 \\
\text{Faraday's law:} \qquad & \vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \\
\text{Generalized Ampere's law:} \qquad & \vec{\nabla} \times \vec{B} = \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t}\ .
\end{aligned}
$$

These four equations *completely* summarize 8.022. Every phenomenon in electricity and magnetism can be derived from these equations. Many of our most important tools for various analyses come from the *integral* version of these equations, which we will discuss shortly.

19.3 The displacement current

What does this new term mean? It turns out that it actually has a very nice, physical interpretation. We can begin to understand this using the old trick of dimensional analysis: electric field has the units of (charge)/(length)^2. The rate of change of electric field has units of (charge)/[(time)(length)^2]. This is the same thing as (current)/(length)^2, which is current density. The generalized Ampere's law is thus often written

$$\vec{\nabla} \times \vec{B} = \frac{4\pi}{c}\left(\vec{J} + \vec{J}_d\right)$$

where

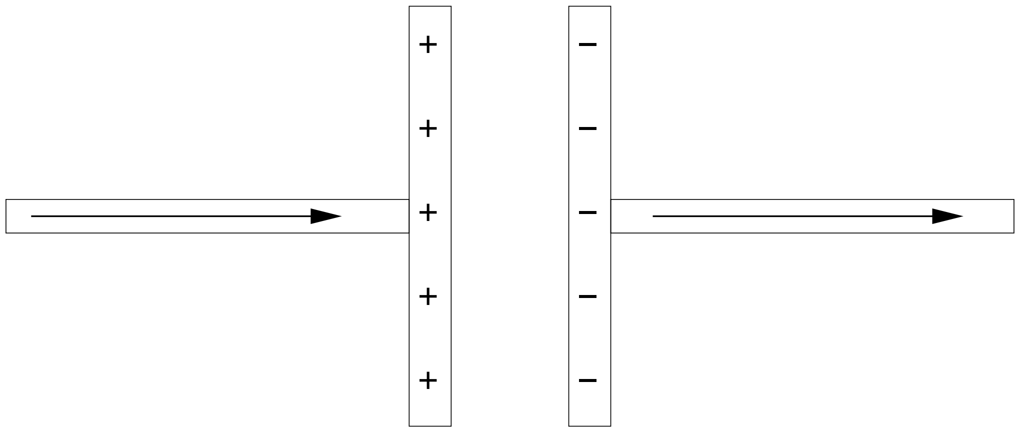
$$\vec{J}_d = \frac{1}{4\pi}\frac{\partial \vec{E}}{\partial t}$$

is known as the *displacement current*, or, more correctly, the displacement current *density*.

The displacement current is not a "real" current, in the sense that it does not describe charges flowing through some region. However, it *acts* just like a real current. Whenever we have a changing $\vec{E}$ field, we can treat its effects as due to the displacement current density arising from that field's variations.

[1]Note that there's a serious typo in Purcell's listing of these equations in SI units, Eq. (15') of Chapter 9.

In particular, the displacement current helps to fix up a few subtleties in circuits that might have bothered you. (I know for a fact that some students have wondered about these issues, and have asked some very good questions!) Consider a current flowing down a wire and charging up a capacitor:

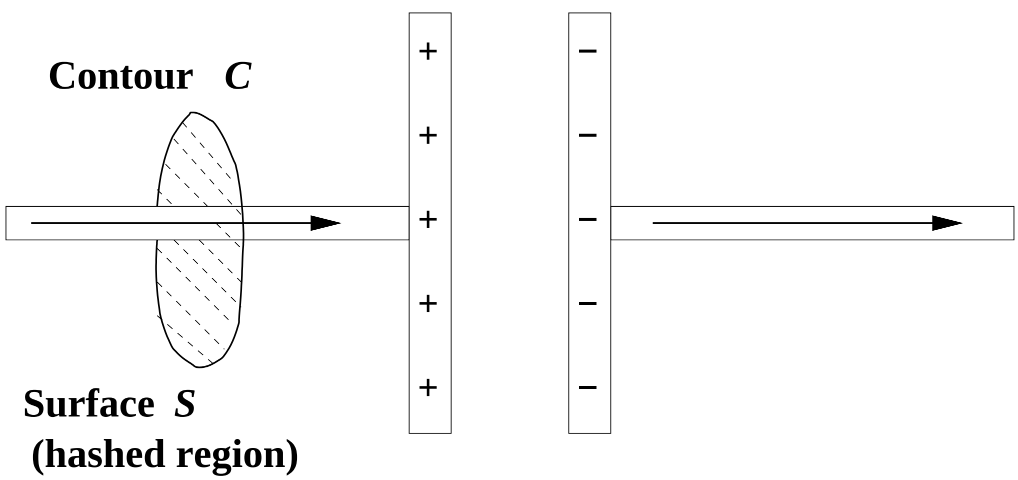


The old integral formulation of Ampere's law (which can be derived from the differential form) tells us

$$\begin{aligned} \oint_C \vec{B} \cdot d\vec{s} &= \frac{4\pi}{c} I_{\rm encl} \\ I_{\rm encl} &= \int_S \vec{J} \cdot d\vec{a} \,. \end{aligned}$$

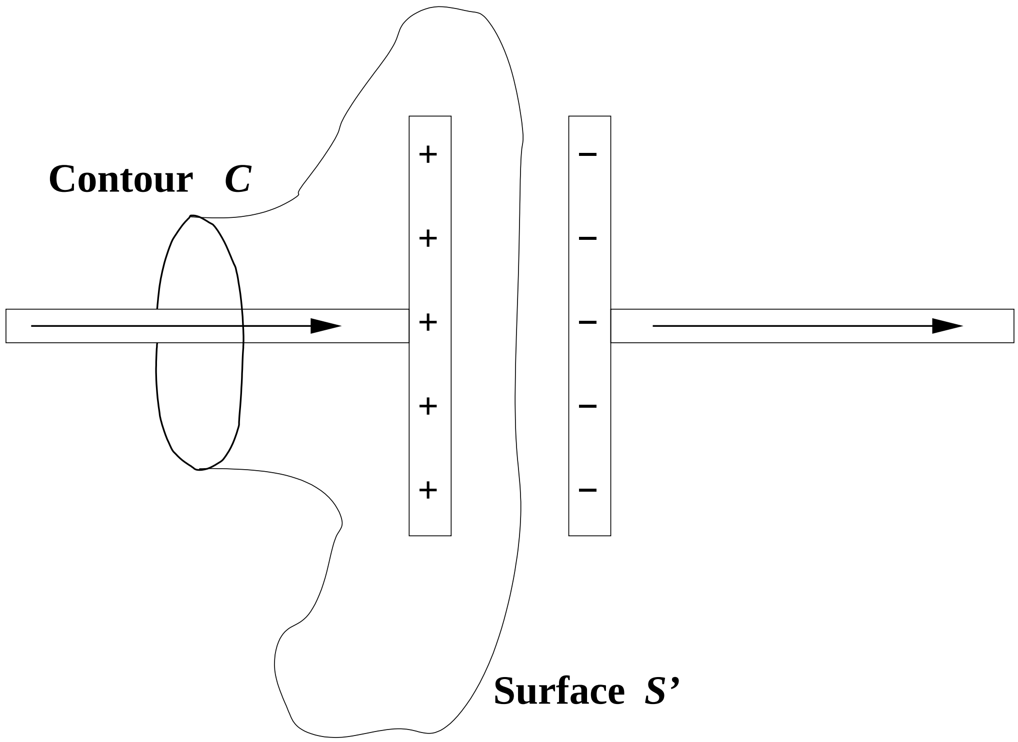
In words, the line integral of the magnetic field around a closed contour C equals $4\pi/c$ times the current enclosed by that contour. The current enclosed by that contour is given by integrating the flux of the current density through a surface S which is bounded by C.

We've normally used this law in the following way: we make a little "Amperian loop" around one of our wires. The contour C is this loop; the surface S is the disk for which S is the border:



The current $I_{\rm encl}$ is then just the current I flowing into the capacitor. Nice, simple, logical.

We now make this complicated. To go between the differential form of Ampere's law and this integral form of it, we use Stoke's theorem. A key fact of using this theorem is that the surface S can be ANY surface that has C as its boundary. In particular, we can choose this surface, S':



The amount of current flowing through this surface is ZERO: no charge flows between the plates! However, there is a flux of *displacement current* between the plates. When we turn our "generalized Ampere's law" into an integral form, we get

$$\begin{aligned}
\oint_C \vec{B}\cdot d\vec{s} &= \frac{4\pi}{c}\left(I_{\rm encl} + I_{d,\ \rm encl}\right) \\
\text{where} \qquad I_{d,\ \rm encl} &= \int_S \vec{J}_d \cdot d\vec{a} \\
&= \frac{1}{4\pi}\int_S \frac{\partial \vec{E}}{\partial t}\cdot d\vec{a} \\
&= \frac{1}{4\pi}\frac{\partial}{\partial t}\int_S \vec{E}\cdot d\vec{a} \\
&= \frac{1}{4\pi}\frac{\partial \Phi_E}{\partial t}\ .
\end{aligned}$$

The displacement current flowing through a surface S (as opposed to the displacement current density) is simply related to the rate of change of electric flux through that surface.

It is simple to show that, for the goofy surface S' drawn above, this rate of change of electric flux is exactly what we need to fix up Ampere's law. First, note that the electric field between the plates points in the same direction as the current; call that direction $\hat{x}$. At any given instant, the electric field between the plates (approximating it to be perfectly uniform) is given by

$$\vec{E} = \frac{4\pi Q}{A}\hat{x}$$

where A is the area of the plates. The electric flux between the plates is

$$\Phi_E = 4\pi Q$$

— not too surprising, given Gauss's law! The rate of change of this flux is

$$\frac{\partial \Phi_E}{\partial t} = 4\pi \frac{\partial Q}{\partial t} = 4\pi I \ .$$

(Since we are charging up this capacitor, $I = dQ/dt = \partial Q/\partial t$. Partial derivatives and total derivatives are the same since the charge only depends on time.) The displacement current "flowing" between the plates is thus

$$I_{d,\ \mathrm{encl}} = \frac{1}{4\pi}\frac{\partial \Phi_E}{\partial t} = I \ .$$

This tells us that

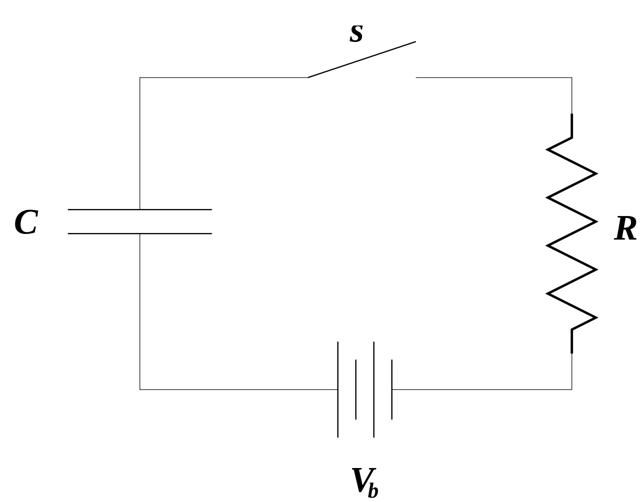
$$\int_S \vec{J} \cdot d\vec{a} = \int_{S'} \vec{J}_d \cdot d\vec{a} = I \ .$$

No matter which surface we use, we find

$$\oint_C \vec{B} \cdot d\vec{s} = \frac{4\pi I}{c} \ .$$

That's good!

Before moving on, note that the displacement current fixes up a somewhat bizarre aspect of circuits with capacitors that I know had some people a tad confused. When we analyze a circuit like this,



we claim that the same current $I(t)$ flows through all the circuit elements. However, we know that this *cannot* be totally true — there is no way for "real" current to make it across the plates of the capacitor. However, *displacement* current certainly makes it across! The displacement current plays the role of continuing[2] the "real" current through the gap in the capacitor. As such, it plays a very important role in ensuring that our use of Kirchhoff's laws is valid.

19.3.1 Integral form of Maxwell's equations

To wrap this section up, let's write down the four Maxwell equations in integral form. These are easily derived by plugging the differential forms into integrals and invoking various vector theorems; hopefully, such manipulations are close to second nature for you now.

$$\text{Gauss's law:} \qquad \Phi_E(S) = \oint_S \vec{E} \cdot d\vec{a} = 4\pi Q_{\rm encl} \; .$$

In words: **The electric flux through a closed surface S equals 4π times the charge enclosed by S.**

$$\text{Magnetic Gauss's law:} \qquad \Phi_B(S) = \oint_S \vec{B} \cdot d\vec{a} = 0 \; .$$

In words: **The magnetic flux through a closed surface S equals 0**. We've never actually derived this, but it follows quite naturally from our earlier discussion of Gauss's law (no magnetic point charges!) and the rule $\vec{\nabla} \cdot \vec{B} = 0$.

$$\text{Faraday's law:} \qquad \mathcal{E} = \oint_C \vec{E} \cdot d\vec{s} = -\frac{1}{c}\frac{\partial \Phi_B}{\partial t} = -\frac{1}{c}\frac{\partial}{\partial t}\int_S \vec{B} \cdot d\vec{a} \; .$$

In words: **The EMF induced around a closed contour C equals $-1/c$ times the rate of change of the magnetic flux through the surface S bounded by C.**

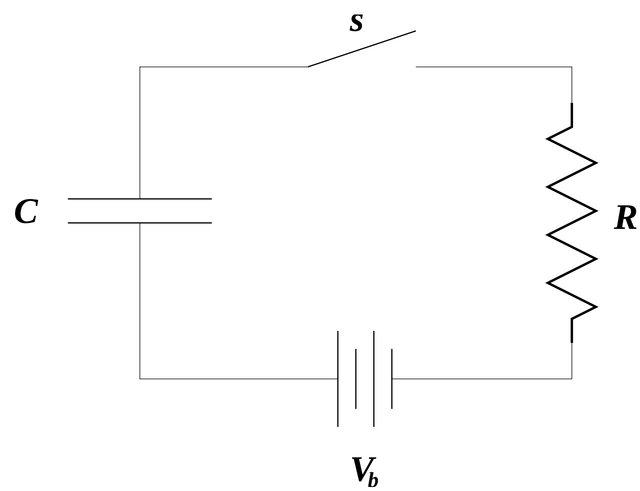
$$\begin{aligned} \text{Generalized Ampere's law:} \qquad & \oint_C \vec{B} \cdot d\vec{s} = \frac{4\pi}{c}\left(I + I_d\right) \\ \text{where} \qquad & I = \int_S \vec{J} \cdot d\vec{a} \\ & I_d = \int_S \vec{J}_d \cdot d\vec{a} = \frac{1}{4\pi}\frac{\partial \Phi_E(S)}{\partial t} \; . \end{aligned}$$

In words: **The line integral of the magnetic field around a closed contour C equals $4\pi/c$ times the sum of the total current — real plus displacement — passing through that contour**.

[2] Indeed, in many older textbooks, the phrase "continuation current" is found instead of displacement current, reflecting the fact it continues the real current.

19.4 Example: Using the displacement current

The beautiful thing about the displacement current is that we can now treat it just like a regular current. Consider this example: suppose we have an RC circuit, and we charge up the capacitor:



As the capacitor charges up, a displacement current flows between the plates. As a consequence, a *magnetic* field is induced between the plates, in addition to the electric field that is building up. We'll now calculate this magnetic field.

The key thing we need is the electric field between the plates as a function of time. To keep things simple, let's assume the plates are circular, with radius a, and we'll assume that they are close enough that we can ignore edge effects. Then, the electric field between the plates as a function of time is

$$E(t) = \frac{4\pi Q(t)}{\pi a^2} \; .$$

The displacement current *density* is

$$\begin{aligned} J_d(t) &= \frac{1}{4\pi}\frac{\partial E}{\partial t} = \frac{1}{\pi a^2}\frac{\partial Q}{\partial t} \\ &= \frac{I(t)}{\pi a^2} \; . \end{aligned}$$

Note that this is spatially constant, though it varies with time.

This circuit is something we beat to death long ago, so we can just quote the results for the current we found then:

$$I(t) = \frac{V_b}{R} e^{-t/RC} \; .$$

We're now ready to compute magnetic fields inside the plates. Since no "real" current flows there, our generalized Ampere's law reduces to

$$\oint_C \vec{B} \cdot d\vec{s} = \frac{4\pi}{c} \int_S \vec{J}_d \cdot d\vec{a} \; .$$

Let's pick as our contour C a circular path of radius $r < a$; the simplest surface S is then the disk of radius r. Then,

$$\oint_C \vec{B} \cdot d\vec{s} = 2\pi r B(r) \ .$$

For the right hand side, we need

$$\frac{4\pi}{c} \int_S \vec{J}_d \cdot d\vec{a} = \frac{4\pi I(t) r^2}{ca^2} \ .$$

The magnetic field is thus

$$B(r) = \frac{2rI(t)}{ca^2} = \frac{2rV_b}{cRa^2} e^{-t/RC} \ .$$

Direction is of course given by right-hand rule: point your right thumb parallel to $\vec{J}_d$ and your fingers curl in the same sense $\vec{B}$.

19.5 Source-free Maxwell's equations

As we shall see shortly, an extremely important limit of Maxwell's equations is found when there are no sources: $\rho = 0$, $\vec{J} = 0$. The equations become

$$\begin{aligned}
\vec{\nabla} \cdot \vec{E} &= 0 \\
\vec{\nabla} \cdot \vec{B} &= 0 \\
\vec{\nabla} \times \vec{E} &= -\frac{1}{c} \frac{\partial \vec{B}}{\partial t} \\
\vec{\nabla} \times \vec{B} &= \frac{1}{c} \frac{\partial \vec{E}}{\partial t} \ .
\end{aligned}$$

These equations are almost perfectly symmetric! Indeed, when Maxwell first introduced the new term in the "generalized Ampere's law", this was his major motivation: he felt that there must be an underlying symmetry between electric and magnetic fields. It is a mark of his remarkable insight that this feeling was entirely correct.

Even more significantly, these equations point to an extremely interesting and important coupling between the magnetic and electric fields. If we imagine we have a time varying electric field, this shows that it will "source" some kind of magnetic field. That magnetic field will be time varying, and so it will source an electric field. That electric field will then source a magnetic field, which will source an electric field, which will ...

You get the point. We end up with an infinite chain of electric to magnetic to electric to ... This is *radiation*. It will be our next major topic.

Scott Hughes *28 April 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 20:
WAVE EQUATION & ELECTROMAGNETIC RADIATION

20.1 Revisiting Maxwell's equations

In our last lecture, we finally ended up with Maxwell's equations, the four equations which encapsulate everything we know about electricity and magnetism. These equations are:

$$\begin{aligned}
\text{Gauss's law:} &\qquad \vec\nabla\cdot\vec E = 4\pi\rho\\
\text{Magnetic law:} &\qquad \vec\nabla\cdot\vec B = 0\\
\text{Faraday's law:} &\qquad \vec\nabla\times\vec E = -\frac{1}{c}\frac{\partial\vec B}{\partial t}\\
\text{Generalized Ampere's law:} &\qquad \vec\nabla\times\vec B = \frac{4\pi}{c}\vec J + \frac{1}{c}\frac{\partial\vec E}{\partial t}\ .
\end{aligned}$$

In this lecture, we will focus on the source free versions of these equations: we set $\rho = 0$ and $\vec J = 0$. We then have

$$\begin{aligned}
\vec\nabla\cdot\vec E &= 0\\
\vec\nabla\cdot\vec B &= 0\\
\vec\nabla\times\vec E &= -\frac{1}{c}\frac{\partial\vec B}{\partial t}\\
\vec\nabla\times\vec B &= \frac{1}{c}\frac{\partial\vec E}{\partial t}\ .
\end{aligned}$$

The source free Maxwell's equations show us that $\vec E$ and $\vec B$ are *coupled*: variations in $\vec E$ act as a source for $\vec B$, which in turn acts as a source for $\vec E$, which in turn acts as a source for $\vec B$, which ... The goal of this lecture is to fully understand this coupled behavior.

To do so, we will find it easiest to first *uncouple* these equations. We do this by taking the curl of each equation. Let's begin by looking at

$$\vec\nabla\times\left(\vec\nabla\times\vec E\right) = \vec\nabla\times\left(-\frac{1}{c}\frac{\partial\vec B}{\partial t}\right)\ .$$

The curl of the left-hand side of this equation is

$$\vec\nabla\times\vec\nabla\times\vec E = \vec\nabla\left(\vec\nabla\cdot E\right) - \nabla^2\vec E = -\nabla^2\vec E\ .$$

(We used this curl identity back in Lecture 13; it is simple to prove, albeit not exactly something you'd want to do at a party.) The simplification follows because we have restricted ourselves to the source free equations — we have $\vec\nabla\cdot\vec E = 0$. Now, look at the curl of the right-hand side:

$$\vec\nabla\times\left(-\frac{1}{c}\frac{\partial\vec B}{\partial t}\right) = -\frac{1}{c}\frac{\partial}{\partial t}\left(\vec\nabla\times\vec B\right) = -\frac{1}{c^2}\frac{\partial^2\vec E}{\partial t^2}\ .$$

Putting the left and right sides together, we end up with

$$\frac{\partial^2 \vec{E}}{\partial t^2} - c^2 \nabla^2 \vec{E} = 0 .$$

Repeating this procedure for the other equation, we end up with something that is essentially identical, but for the magnetic field:

$$\frac{\partial^2 \vec{B}}{\partial t^2} - c^2 \nabla^2 \vec{B} = 0 .$$

As we shall now discuss, these equations are particularly special and important.

20.2 The wave equation

Let us focus on the equation for $\vec{E}$; everything we do will obviously pertain to the $\vec{B}$ equation as well. Furthermore, we will simplify things initially by imagining that $\vec{E}$ only depends on x and t. The equation we derived for $\vec{E}$ then reduces to

$$\frac{\partial^2 \vec{E}}{\partial t^2} - c^2 \frac{\partial^2 \vec{E}}{\partial x^2} = 0 .$$

20.2.1 General considerations

At this point, it is worth taking a brief detour to talk about equations of this form more generally. This equation for the electric field is a special case of

$$\frac{\partial^2 f}{\partial t^2} - v^2 \frac{\partial^2 f}{\partial x^2} = 0 .$$

This equation is satisfied by *ANY* function whatsoever[1] provided that the argument of the function is written in the following special way:

$$f = f(x \pm vt) .$$

This is easy to prove. Let $u = x \pm vt$, so that $f = f(u)$. Then, using the chain rule,

$$\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x}\frac{\partial f}{\partial u} = \frac{\partial f}{\partial u} ;$$

it follows quite obviously that

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial u^2} .$$

The time derivatives are not quite so trivial:

$$\frac{\partial f}{\partial t} = \frac{\partial u}{\partial t}\frac{\partial f}{\partial u} = \pm v \frac{\partial f}{\partial u} ,$$

[1] To the mathematical purists, we must insist that the function f be smooth enough that it has well-behaved derivatives. At the 8.022 level, this is a detail we can largely ignore.

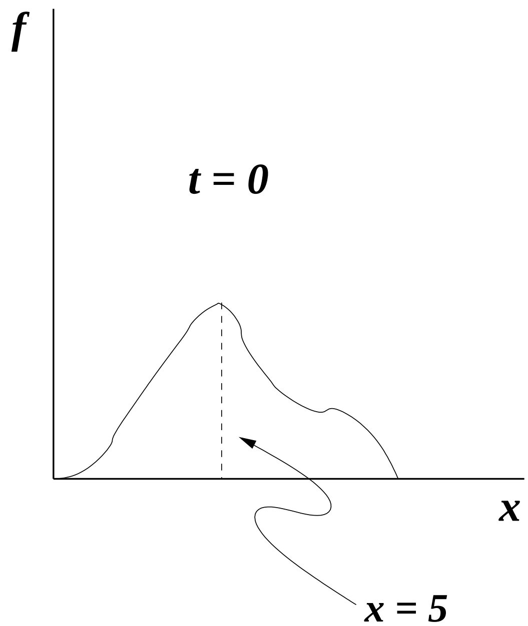
so

$$\frac{\partial^2 f}{\partial t^2} = v^2 \frac{\partial^2 f}{\partial u^2} .$$

Plugging in, we see

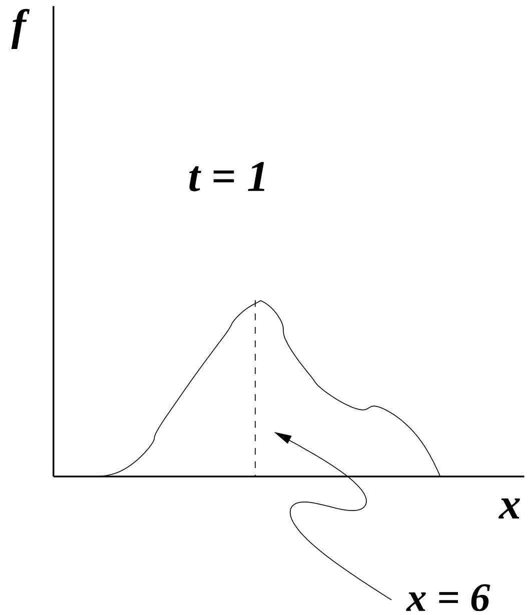
$$\frac{\partial^2 f}{\partial t^2} - v^2 \frac{\partial^2 f}{\partial x^2} = v^2 \frac{\partial^2 f}{\partial u^2} - v^2 \frac{\partial^2 f}{\partial u^2} = 0 .$$

Having shown that any function of the form $f(x \pm vt)$ satisfies this equation, we now need to examine what this solution means. Let's consider $f(x - vt)$ first; and, let's say that $v = 1$. (Don't worry about units just yet.) Suppose that at $t = 0$, the function f has the following shape:



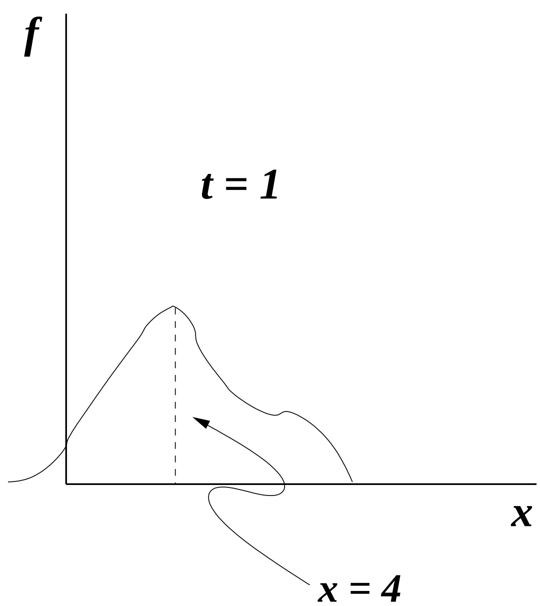
The peak of this function occurs when the argument of f is 5 — in other words, for $t = 0$, the peak occurs at $x = 5$.

What does this function look at $t = 1$? Clearly it will have the exact same shape. *However*, that shape will be shifted over on the x-axis by one unit! For example, the peak of f still occurs when the argument of f is 5. This means we must have $x - vt = 5$, which for $t = 1$ means the peak occurs at $x = 6$. All other points in the shape are similarly shifted:



As time passes, this shape just slides to the right, in the direction of increasing x. In other words, $f(x - vt)$ *representes a wave traveling in the POSITIVE* x *direction with speed* v.

What is the other solution? Applying this logic to $f(x + vt)$, we see that this represents a wave moving in the opposite direction, towards smaller x. For example, at $t = 1$, the $x + vt = 5$ means $x = 4$ — the peak of the curve shifts to $x = 4$ after 1 time unit:



$f(x + vt)$ *representes a wave traveling in the NEGATIVE* x *direction with speed* v.

OK, what we've got so far is that equations of the form

$$\frac{\partial^2 f}{\partial t^2} - v^2 \frac{\partial^2 f}{\partial x^2} = 0$$

generically have solutions of the form $f = f(x \pm vt)$, where $f(u)$ is essentially arbitrary. These solutions represent waves travelling at speed v; the solution with $x - vt$ represents a wave travelling in the positive x direction; the solution with $x + vt$ represents a wave travelling in the negative x direction.

20.2.2 Maxwell and the wave equation

You can probably see where we're going with this now! The equation for $\vec{E}$ that we derived earlier,

$$\frac{\partial^2 \vec{E}}{\partial t^2} - c^2 \frac{\partial^2 \vec{E}}{\partial x^2} = 0$$

is itself a wave equation. It represents a wave travelling along the x axis with speed c — the speed of light! This kind of analysis is what made people realize that light is itself an electromagnetic wave.

This would have been far more interesting if we had worked in SI units throughout — getting the speed of light at the end of the day in cgs units is one of those "Well, duh" moments, since c appears throughout those equations already. In SI units, the wave equation that we end up with is

$$\frac{\partial^2 \vec{E}}{\partial t^2} - \frac{1}{\epsilon_0 \mu_0} \frac{\partial^2 \vec{E}}{\partial x^2} = 0$$

indicating that electromagnetic waves travel with speed

$$v = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \ .$$

Plug in some numbers:

$$\begin{aligned} \epsilon_0 &= 8.85418 \times 10^{-12}\,\text{Coulomb}^2\ \text{Newton}^{-1}\ \text{meter}^{-2} \\ \mu_0 &= 4\pi \times 10^{-7}\,\text{Newton sec}^2\ \text{Coulomb}^{-2} \end{aligned}$$

so,

$$\epsilon_0 \mu_0 = 1.11265 \times 10^{-17}\,\text{s}^2\ \text{m}^{-2}$$

$$\longrightarrow \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 2.998 \times 10^8\,\text{m/s} = 2.998 \times 10^{10}\,\text{cm/s} \ .$$

This again is of course the speed of light.

The fact that light is connected to electric and magnetic fields was first recognized from this kind of calculation. Maxwell first showed that the four equations describing electric and magnetic fields could be rearranged to form a wave equation, with a speed of propagation as given above.

At the time, ϵ_0 and μ_0 were empirical quantities that were determined through careful measurements of static electric and magnetic fields — in SI units, ϵ_0 shows up in the capacitance formula, and μ_0 shows up in the inductance formula . It was quite a shock to find that these quantities determined the behavior of highly *dynamical* fields as well, and even more of a shock to realize that the speed $1/\sqrt{\epsilon_0\mu_0}$ was the speed of light! At the time, ϵ_0, μ_0, and c were uncertain enough that the correspondance between $1/\sqrt{\epsilon_0\mu_0}$ was not perfectly definitive. However, it was close enough that Maxwell wrote in 1864[2]

This velocity is so nearly that of light that it seems we have strong reason to conclude that light itself (including radiant heat and other radiations) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.

It is largely for this reason that this whole set of equations is named for Maxwell.

20.3 Plane waves

A particularly important and useful form of wave for our study are the class known as *plane waves*. The most general form of these waves is

$$\begin{aligned}
\vec{E} &= \vec{E}_0 \sin(\vec{k}\cdot\vec{r} - \omega t) \\
&= \vec{E}_0 \sin(k_x x + k_y y + k_z z - \omega t) \ , \\
\vec{B} &= \vec{B}_0 \sin(\vec{k}\cdot\vec{r} - \omega t) \\
&= \vec{B}_0 \sin(k_x x + k_y y + k_z z - \omega t) \ .
\end{aligned}$$

The quantities $\vec{E}_0$ and $\vec{B}_0$ are the amplitudes of the electric and magnetic fields. Note that these amplitudes are vectors; we will determine some important properties for these vectors soon.

The vector $\vec{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$ is the "wavevector"; its magnitude $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ is the "wavenumber". The unit vector $\hat{k} = \vec{k}/k$ is the wave's propagation direction. The number ω is the angular frequency of the wave.

Earlier, we showed that electromagnetic waves are given by functions of the form $f(x-ct)$. Generalizing to three dimensions, we would write this as $f(\vec{r} - c\hat{k}t)$ — the position $\vec{r}$ changes with velocity $c\hat{k}$. With this in mind, we can quickly show that the wavenumber k and ω are related to one another:

$$\vec{k}\cdot\vec{r} - \omega t = \vec{k}\cdot\left(\vec{r} - \frac{\omega}{k}\hat{k}t\right) \ .$$

Combining this with the requirement that the dependence on $\vec{r}$ and t must be of the form $f(\vec{r} - c\hat{k}t)$ tells us that k and ω are related by the speed of light:

$$\omega = ck \ .$$

[2]See R. Baierlein, Newton to Einstein (Cambridge, 1992).

20.3.1 More on $\vec{k}$ and ω

For the moment, let's simplify our wave vector so that it is oriented along the x-axis: we put $\vec{k} = k\hat{x}$. Our plane wave solution is then

$$\begin{aligned}
\vec{E} &= \vec{E}_0 \sin(kx - \omega t) \\
\vec{B} &= \vec{B}_0 \sin(kx - \omega t) \; .
\end{aligned}$$

Let's look at the spatial variation of this wave at $t = 0$: clearly, we just have

$$\begin{aligned}
\vec{E} &= \vec{E}_0 \sin(kx) \\
\vec{B} &= \vec{B}_0 \sin(kx) \; .
\end{aligned}$$

These functions pass through zero at $x = 0$, $x = 2\pi/k$, $x = 4\pi/k$, etc. $2\pi/k$ is clearly a special lengthscale: it is called the *wavelength*, λ. It tells us about the lengthscale over which the wave repeats its pattern.

Now, think about the time variations at a particular location, say $x = 0$:

$$\begin{aligned}
\vec{E} &= -\vec{E}_0 \sin(\omega t) \\
\vec{B} &= -\vec{B}_0 \sin(\omega t) \; .
\end{aligned}$$

These functions pass through zero at $t = 0$, $t = 2\pi/\omega$, $t = 4\pi/\omega$, etc. $2\pi/\omega$ is the *period* of the wave, T. This of course the usual relationship between the period of an oscillation and its angular frequency. The inverse period is the wave's frequency (as opposed to *angular* frequency), usually written ν or f:

$$\nu = f = \omega/2\pi = 1/T \; .$$

20.3.2 Summary of wavenumber and frequency relations

Putting all of this together we end up with several relations between wavenumber, wavelength, frequency, angular frequency, and the speed of light. First, we have

$$\omega = ck \; .$$

Plugging in $\omega = 2\pi\nu$ and $k = 2\pi/\lambda$, we find

$$\nu\lambda = c \; .$$

20.4 Final constraints

Let's assess what we have so far: the source free Maxwell equations,

$$\begin{aligned}
\vec{\nabla} \cdot \vec{E} &= 0 \\
\vec{\nabla} \cdot \vec{B} &= 0 \\
\vec{\nabla} \times \vec{E} &= -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \\
\vec{\nabla} \times \vec{B} &= \frac{1}{c}\frac{\partial \vec{E}}{\partial t} \; ,
\end{aligned}$$

can be manipulated in such a way as to show that $\vec{E}$ and $\vec{B}$ each obey the wave equation,

$$
\begin{aligned}
\frac{\partial^2\vec{E}}{\partial t^2} - c^2\nabla^2\vec{E} &= 0\\
\frac{\partial^2\vec{B}}{\partial t^2} - c^2\nabla^2\vec{B} &= 0\ .
\end{aligned}
$$

We're not done yet! In addition to satisfying the wave equation, we must also verify that our functions $\vec{E}$ and $\vec{B}$ satisfy the *other* Maxwell equations — I can write down *lots* of functions that satisfy the wave equations, but that are not valid electromagnetic fields for $\rho = 0$ and $\vec{J} = 0$.

First, $\vec{E}$ and $\vec{B}$ must satisfy $\vec{\nabla}\cdot\vec{E} = 0$ and $\vec{\nabla}\cdot\vec{B} = 0$. Let's plug our plane wave solutions into this:

$$
\begin{aligned}
\vec{\nabla}\cdot\vec{E} &= \vec{\nabla}\cdot\left[\vec{E}_0\sin(k_x x + k_y y + k_z z - \omega t)\right]\\
&= (k_x E_{0,x} + k_y E_{0,y} + k_z E_{0,z})\cos(k_x x + k_y y + k_z z - \omega t)\\
&= \left(\vec{k}\cdot\vec{E}_0\right)\cos(k_x x + k_y y + k_z z - \omega t)\ .
\end{aligned}
$$

Likewise,

$$
\vec{\nabla}\cdot\vec{B} = \left(\vec{k}\cdot\vec{B}_0\right)\cos(k_x x + k_y y + k_z z - \omega t)\ .
$$

The only way for this to hold for *all* x and t is if

$$
\begin{aligned}
\vec{k}\cdot\vec{E}_0 &= 0\\
\vec{k}\cdot\vec{B}_0 &= 0\ .
\end{aligned}
$$

In other words, the divergence equations require that

Radiation's electric and magnetic fields are orthogonal to the propagation direction.

Let's next look at one of the curl equations:

$$
\vec{\nabla}\times\vec{B} = \frac{1}{c}\frac{\partial\vec{E}}{\partial t}\ .
$$

This equation tells us that $\vec{E}$ and $\vec{B}$ are perpendicular to one another. On the right-hand side, we have $\partial\vec{E}/\partial t$. This points in the same direction as $\vec{E}$ — the time derivative does not change the vector's direction. On the left, we have $\vec{\nabla}\times\vec{B}$. This is perpendicular to $\vec{B}$ — the curl of any vector is orthogonal to that vector.

So, the first thing the curl equations tell us is that

Radiation's electric and magnetic fields are orthogonal to each other.

Now, let's actually compute the derivatives. The right-hand side is easiest, so let's do it first:

$$
\begin{aligned}
\frac{1}{c}\frac{\partial\vec{E}}{\partial t} &= -\frac{\omega}{c}\vec{E}_0\cos(\vec{k}\cdot\vec{r} - \omega t)\\
&= -k\vec{E}_0\cos(\vec{k}\cdot\vec{r} - \omega t)\ .
\end{aligned}
$$

Now, the left-hand side:

$$\begin{aligned}\vec{\nabla} \times \vec{B} &= \vec{\nabla} \times \left[\vec{B}_0 \sin\left(\vec{k} \cdot \vec{r} - \omega t\right)\right] \\ &= \vec{\nabla} \times \left[\vec{B}_0 \sin\left(k_x x + k_y y + k_z z - \omega t\right)\right] \ .\end{aligned}$$

This looks like it might be a mess. However, back on Pset 3 we proved a very useful identity:

$$\vec{\nabla} \times (\vec{F} f) = f \vec{\nabla} \times \vec{F} + \vec{\nabla} f \times \vec{F} \ .$$

Put $\vec{F} \equiv \vec{B}_0$, $f = \sin\left(\vec{k} \cdot \vec{r} - \omega t\right)$. Then, we have $\vec{\nabla} \times \vec{F} = 0$ — $\vec{B}_0$ is constant! Only the other term contributes; plugging in for $\vec{F}$ and f, what we are left with is

$$\begin{aligned}\vec{\nabla} \times \vec{B} &= \vec{\nabla}\left[\sin\left(k_x x + k_y y + k_z z - \omega t\right)\right] \times \vec{B}_0 \\ &= \left[\left(k_x \hat{x} + k_y \hat{y} + k_z \hat{z}\right) \cos\left(k_x x + k_y y + k_z z - \omega t\right)\right] \times \vec{B}_0 \\ &= \left(\vec{k} \times \vec{B}_0\right) \cos\left(\vec{k} \cdot \vec{r} - \omega t\right) \ .\end{aligned}$$

After slogging through all of this, we finally obtain

$$\begin{aligned}\vec{\nabla} \times \vec{B} &= \frac{1}{c} \frac{\partial \vec{E}}{\partial t} \\ \longrightarrow \left(\vec{k} \times \vec{B}_0\right) \cos\left(\vec{k} \cdot \vec{r} - \omega t\right) &= -k \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t) \ ,\end{aligned}$$

or

$$\vec{E}_0 = -\hat{k} \times \vec{B}_0 \ .$$

There are two interesting things we can do with this. First, let's take the magnitude of both sides:

$$\begin{aligned}|\vec{E}_0| &= |-\hat{k} \times \vec{B}_0| \\ &= |\hat{k}||\vec{B}_0| \qquad \text{since } \hat{k} \text{ is orthogonal to } \vec{B}_0 \\ &= |\vec{B}_0| \qquad \text{since } |\hat{k}| = 1.\end{aligned}$$

In cgs units, the radiation's electric and magnetic fields have the same magnitude.

Second, one can show from this equation that $\vec{E}_0 \times \vec{B}_0 = |\vec{E}_0|^2 \hat{k}$:

$$\begin{aligned}\vec{E}_0 \times \vec{B}_0 &= -(\hat{k} \times \vec{B}_0) \times \vec{B}_0 \\ &= \vec{B}_0 \times (\hat{k} \times \vec{B}_0) \\ &= \hat{k}\left(\vec{B}_0 \cdot \vec{B}_0\right) - \vec{B}_0\left(\hat{k} \cdot \vec{B}_0\right) \\ &= \hat{k}|\vec{E}_0|^2 \ .\end{aligned}$$

To go from the 1st line to the 2nd line, I used $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$; to go to the 3rd line I used the vector identity $\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$. I won't prove this identity explicitly; it's simple (albeit fairly tedious) to prove this in Cartesian coordinates. Finally, to go to the last line I used the results we just proved about the magnitude of the $\vec{B}$ field and its orthogonality to $\hat{k}$. The punchline is that

$\vec{E}_0 \times \vec{B}_0$ is parallel to the propagation direction $\hat{k}$.

As we shall soon see, $\vec{E} \times \vec{B}$ has a separate, *very* important physical meaning: Up to a constant, $\vec{E} \times \vec{B}$ tells us about the flow of energy that is carried by the radiation. This makes it very clear that $\vec{E} \times \vec{B}$ must be parallel to $\hat{k}$!

Scott Hughes *3 May 2005*

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 21:
Polarization & Scattering

21.1 Summary: radiation so far

In the last few lectures, we examined solutions of the source free Maxwell equations:

$$\vec{\nabla}\cdot\vec{E} = 0 \qquad \vec{\nabla}\times\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t}$$
$$\vec{\nabla}\cdot\vec{B} = 0 \qquad \vec{\nabla}\times\vec{B} = \frac{1}{c}\frac{\partial\vec{E}}{\partial t}\ .$$

With a little massaging, we discovered that these equations can be rewritten as *wave equations* for $\vec{E}$ and $\vec{B}$:

$$\frac{\partial^2\vec{E}}{\partial t^2} - c^2\nabla^2\vec{E} = 0$$
$$\frac{\partial^2\vec{B}}{\partial t^2} - c^2\nabla^2\vec{B} = 0\ .$$

A particularly instructive solution to the wave equations are the *plane wave* forms:

$$\vec{E}(\vec{r},t) = \vec{E}_0\sin(\vec{k}\cdot\vec{r} - \omega t)$$
$$\vec{B}(\vec{r},t) = \vec{B}_0\sin(\vec{k}\cdot\vec{r} - \omega t)\ .$$

This solution represents an electromagnetic wave propagating in the $\hat{k} = \vec{k}/k$ direction (where $k = \sqrt{\vec{k}\cdot\vec{k}} = \sqrt{k_x^2 + k_y^2 + k_z^2}$).

By considering how the wave behaves at some fixed time, we learned that k is simply related to the wavelength λ:

$$k = 2\pi/\lambda\ .$$

The requirement that this solution satisfy the wave equation tells us that

$$\omega = ck\ .$$

From the definition $\omega = 2\pi\nu$ (angular frequency is 2π radians times "regular" frequency), we then obtain

$$\lambda\nu = c\ .$$

Finally, requiring that the plane wave solution satisfy all of Maxwell's equations leads to some important constraints on the vector amplitudes $\vec{E}_0$ and $\vec{B}_0$. These constraints are:

- The amplitudes are orthogonal to the propagation direction: $\hat{k} \cdot \vec{E}_0 = 0$, $\hat{k} \cdot \vec{B}_0 = 0$.
- The amplitudes are orthogonal to each otherL $\vec{E}_0 \cdot \vec{B}_0 = 0$.
- The amplitudes have the same magnitude: $|\vec{E}_0| = |\vec{B}_0|$.
- The propagation direction is parallel to $\vec{E} \times \vec{B}$.

These are important and rather constraining conditions. Nonetheless, they leave us with a great deal of freedom in the amplitudes. This freedom is described in terms of the radiation's *polarization state*.

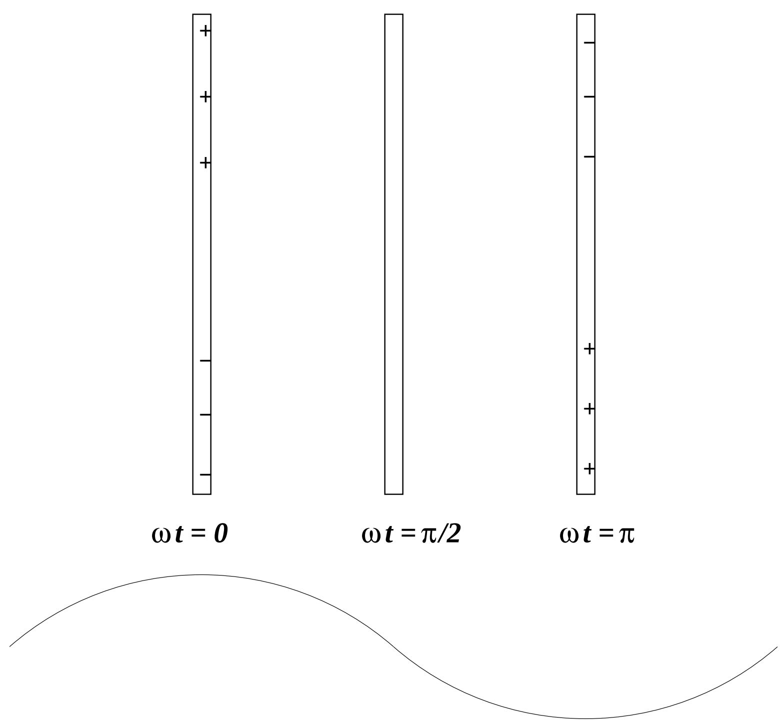
21.2 Linear polarization

Since plane waves propagate in a straight line, we might as well just define their propagation direction as something simple and be done with it. In what follows, we will take $\hat{k} = \hat{z}$, so our wave is of the form

$$\vec{E} = \vec{E}_0 \sin(kz - \omega t)$$
$$\vec{B} = \vec{B}_0 \sin(kz - \omega t)\ .$$

Suppose that the wave is arranged (somehow) so that the electric field is aligned with the x axis: $\vec{E}_0 = E_0\hat{x}$. Then, our requirement that $\vec{E} \times \vec{B}$ be parallel to $\hat{k}$ tells us that $\vec{B}_0 = E_0\hat{y}$. (We're also using $|\vec{E}_0| = |\vec{B}_0|$.)

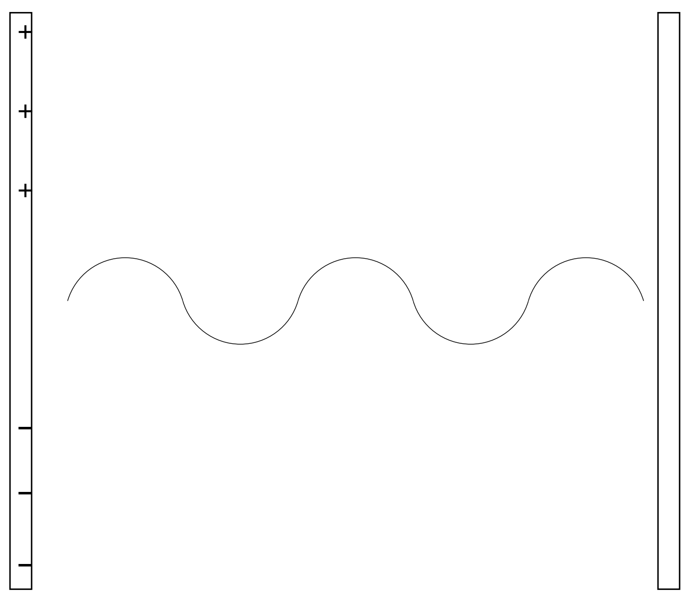
This configuration is known as a *linearly polarized* wave, as the electric (and magnetic) fields at all points align parallel to a line. Such an electromagnetic wave is quite simple to produce: we just need to take a conductor and arrange things so that it has an oscillating charge distribution. Here's an example:



The top part of this figure shows an *antenna*: a long conductor in which we drive a very rapidly oscillating current, so that the conductor has a charge *distribution*. As we move to the right, time increases.

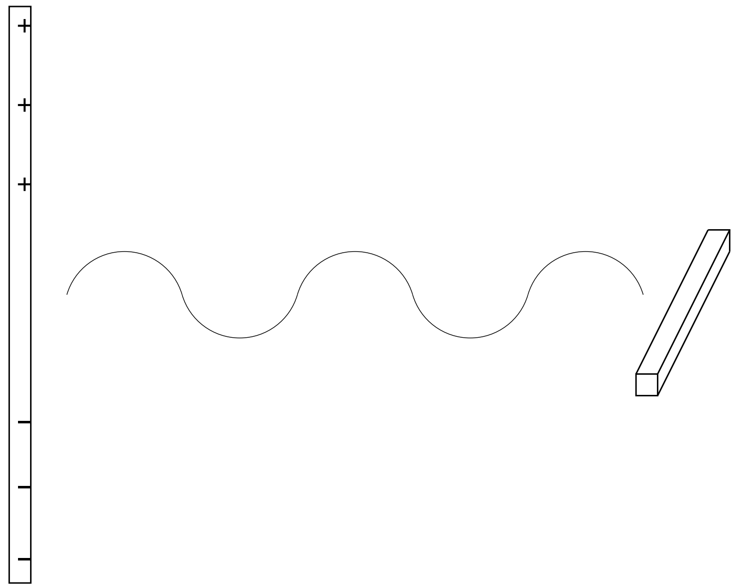
At any given instant, one end is likely to have a net positive charge, the other end a net negative charge; the whole thing is neutral overall. This setup creates an electric field which oscillates in phase with the current's oscillation; this is sketched by the sinusoid below the antenna.

The antenna we have described here is more properly a *generating* or *broadcast* antenna. What if we want to measure the electromagnetic wave that this antenna is broadcasting? Quite simple: we just need to setup a second antenna that will "catch" the oscillating electromagnetic field of the broadcaster. If we put a second antenna down near it, the oscillating electric field will shake charges in this *receiving* antenna:



The oscillating charges themselves produce a current, which we can measure and (possibly) extract information from. This is how a radio works!

When the generated electromagnetic wave is linearly polarized, it is (hopefully) intuitive obvious that the receiving antenna should be oriented parallel to the polarization vector. Suppose we get it wrong by 90°:



For this orientation, the receiving antenna measures doodly squat: there's not enough "room" for a significant charge oscillation to occur.

21.3 Polaroids

Certain substances act like antennas for radiation in the visual band — light that we can see. They typically consist of organic molecules that are fairly extended in one direction. These molecules can carry a current along that direction; but, no current can flow in the direction orthogonal to the molecule's orientation.

A *polaroid* is a sheet of plastic embedded with molecules of this type all lined up along some axis. Suppose linearly polarized light impinges on these molecules. Suppose further that the light is polarized such that $\vec{E}$ is aligned with its organic molecules. Then, the light shakes charges in the polaroid; it generates a current, which generates heat in the plastic sheet. The light is absorbed — nothing gets through.

Suppose instead that light is linearly polarized in a direction *perpendicular* to the sheet's organic molecules. This light will not be able to shake the charges in the sheet — they are not free to move in that direction. Rather than being absorbed, the light passes right through the sheet.

The direction that *allows* light through is, by convention, the "preferred" direction of the polaroid. The punchline of this discussion is that

Polaroids are transparent to light with polarization parallel to their preferred direction. They are opaque to light with polarization perpendicular to their preferred direction.

21.3.1 Changing the polarization direction

What if the light is polarized in a direction somewhere between parallel and perpendicular to the polaroid's preferred orientation? For concreteness, suppose the light's electric field is oriented along the x axis:

$$\vec{E}_{\text{incoming}} = E_0 \hat{x} \cos(kz - \omega t) \ .$$

Suppose the polaroid's orientation $\hat{p}$ is at some angle θ between the x and y axes:

$$\hat{p} = \hat{x} \cos\theta + \hat{y} \sin\theta \ .$$

Since the polaroid has a component of its orientation parallel to the light's electric field, the light will *partially* be able to go through the polaroid. The amount of electric field coming out of the polaroid will be given by the overlap between the incoming electric field and the polaroid's orientation:

$$|\vec{E}_{\text{outgoing}}| = \vec{E}_{\text{incoming}} \cdot \hat{p} = E_0 \cos\theta \cos(kz - \omega t) \ .$$

This outgoing light must be parallel to the polaroid's orientation — the polaroid defines the new polarization direction. Thus, the full solution for the outgoing light's $\vec{E}$ field is

$$\vec{E}_{\text{outgoing}} = \left(\vec{E}_{\text{incoming}} \cdot \hat{p}\right) \hat{p} \ .$$

A polaroid reduces the amplitude of linearly polarization light by $\cos\theta$, where θ is the angle between the polaroid's orientation and the light's electric field. It rotates the orientation of that field by θ.

21.3.2 Random light

In many situations, light is not linearly polarized — the light generated by a bulb, or coming from the surface of the sun, or from a fire, is best described as a superposition of many plane waves, each with their own polarization[1]:

$$\vec{E}_{\text{random}} = \sum_{j=1}^{\text{huge}} E_0 \left(\cos\theta_j \hat{x} + \sin\theta_j \hat{y}\right) \cos(kz - \omega t) \ .$$

We can make this light linearly polarized by passing in through a polaroid. For concreteness, imagine the polaroid is aligned along the x axis, so $\hat{p} = \hat{x}$. The light that comes out of the polaroid is no longer randomly oriented:

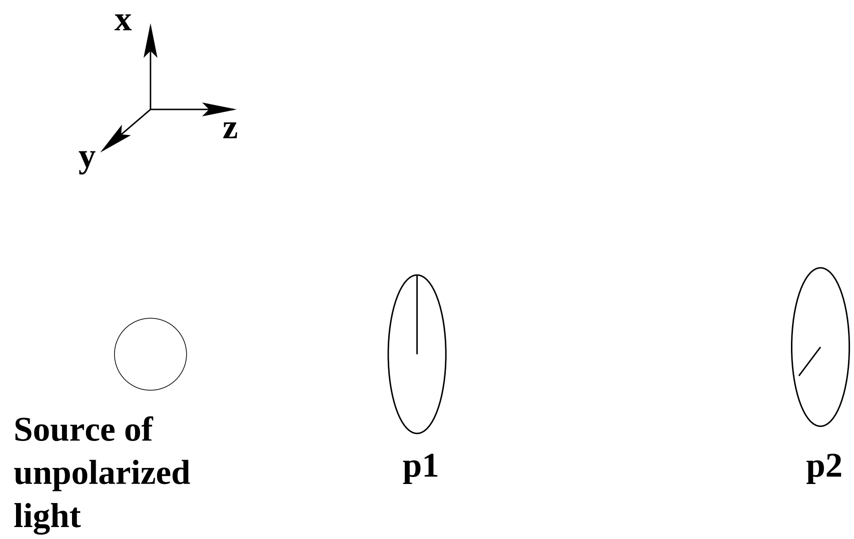
$$\vec{E}_{\text{outgoing}} = E_0 \hat{x} \cos(kz - \omega t) \sum_{j=1}^{\text{huge}} \cos\theta_j \ .$$

Polaroids can thus be used to *produce* linearly polarized light. As we'll discuss in the next lecture, one ends up reducing the intensity (brightness) when this is done; but, the light that comes out has $\vec{E}$ fields all lined up in the same way.

[1] The waves will also in general have an independent phase offset, so that their behavior with respect to time and space should be written as $\cos(kz - \omega t + \varphi_j)$. This complication, though often important, does not play a role in this discussion. An example where it *does* matter, quite a bit, is the laser. Lasers produce light in which all of these "offset phases" line up. This means that the light is *coherent*: the electric and magnetic fields are spatially and temporally in phase with each other.

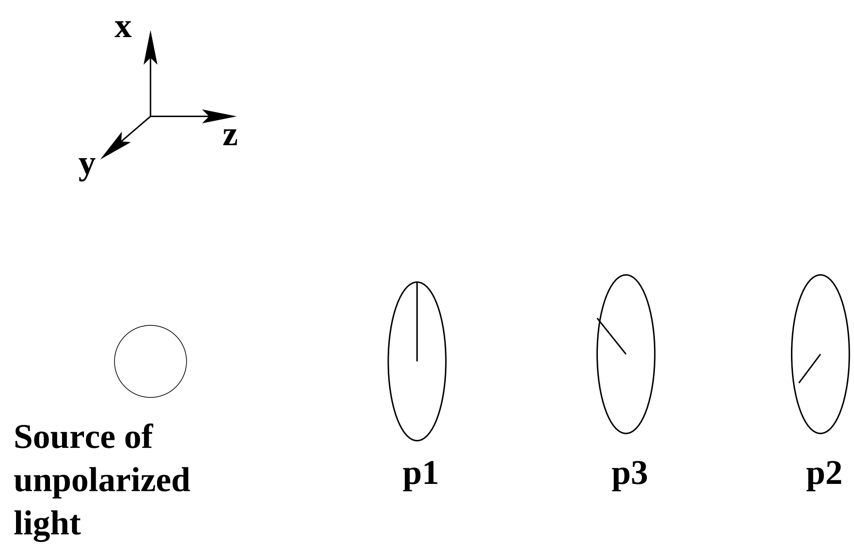
21.3.3 3 polaroids let more light through than 2

From this discussion, we can now easily understand how 3 polaroids let more light through than 2. Consider 2 first: we have one polaroid aligned with the x axis ($\hat{p}_1 = \hat{x}$), and then a second aligned with the y axis ($\hat{p}_2 = \hat{y}$).



After passing through p_1, the light will be polarized along the x axis. It hits p_2, which is aligned with $\hat{y}$. Since $\hat{x}$ and $\hat{y}$ are orthogonal, nothing gets through.

We now place a third polaroid p_3 between p_1 and p_2. We set p_3 to an orientation halfway between p_1 and p_2: $\hat{p}_3 = \cos\theta\hat{x} + \sin\theta\hat{y}$, with $\theta = \pi/4$.

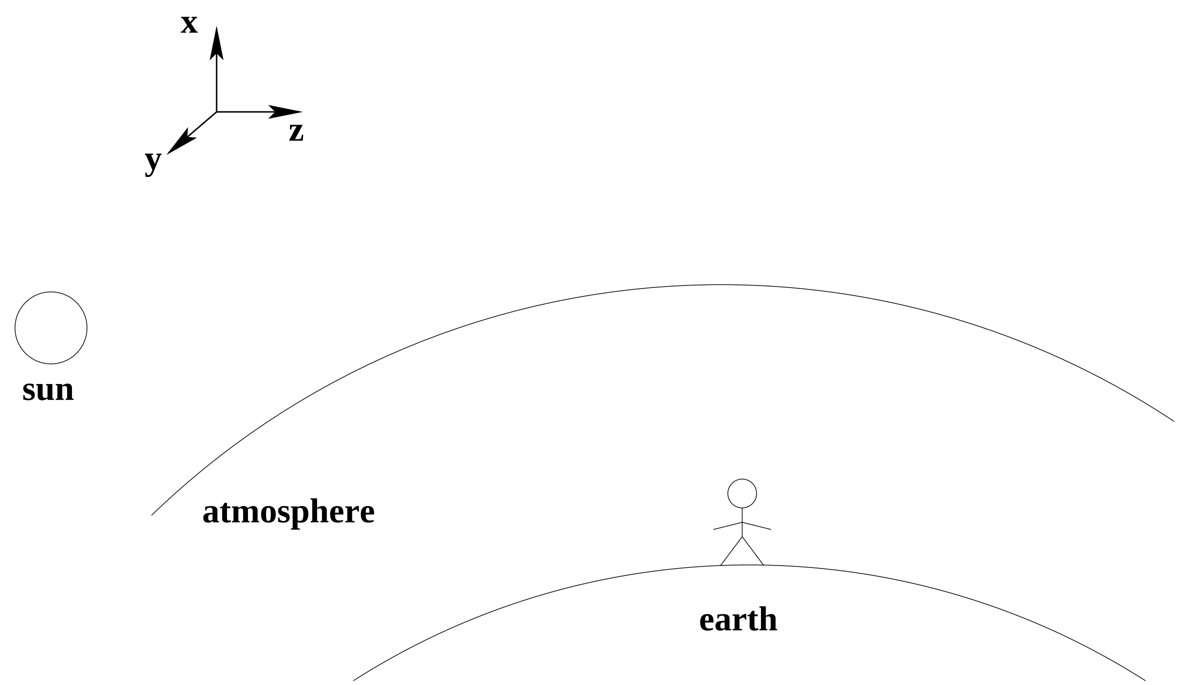


After passing through p_1, the light will be polarized along the x axis. It then hits p_3. The electric fields are reduced by a factor of $\cos\pi/4 = 1/\sqrt{2}$, and the field is then aligned with $\hat{p}_3$. It then hits p_2. The electric fields are *again* reduced by a factor of $\cos\pi/4 = 1/\sqrt{2}$. The fields are *reduced* (by quite a bit!) but they are *not* eliminated.

21.4 Scattering

Many materials *scatter* light: we send light into a medium, and the light is scattered into new directions. This is an extremely rich subject — very dense tomes have been written about the properties and physics of light scattering! Here, we'll content ourselves with a few mostly qualitative observations based on what we've learned so far.

First, consider unpolarized light passing through some medium. The source of unpolarized light could be, for example, the sun; the medium could be the earth's atmosphere.



The (unpolarized) sunlight is initially propagating in the z direction. We look up, in the x direction. What can we say about the light that we see when we look up?

First, we know that, since the light is initially propagating in the z direction, it can have *no* polarization along that axis. Second, since we then "measure" (with our eyes) light that is propagating along the x direction, it can have *no* polarization along that axis. Our conclusion is that **the light that comes down to us from the sky should be linearly polarized along $\hat{y}$**.

In truth, this isn't quite the case. Among other things, the light that we actually measure tends to scatter multiple times. This gives the polarization vectors additional chances to change direction, so we are likely to end up with a mixture rather than a perfectly linear state. But, this argument *does* suggest a general tendency: the light that scatters out of the sky will *tend* to have a preferred polarization axis.

Suppose the light from the sun were somehow made to be linearly polarized. Then, the scattered light will be most intense in the direction *orthogonal* to the polarization direction. If we put a giant polaroid in front of the sun, we could make the sky vary from dark to bright by rotating the polaroid! It would be bright when the polarization axis is orthogonal to "up"; it would be dark when the axis is parallel to "up".

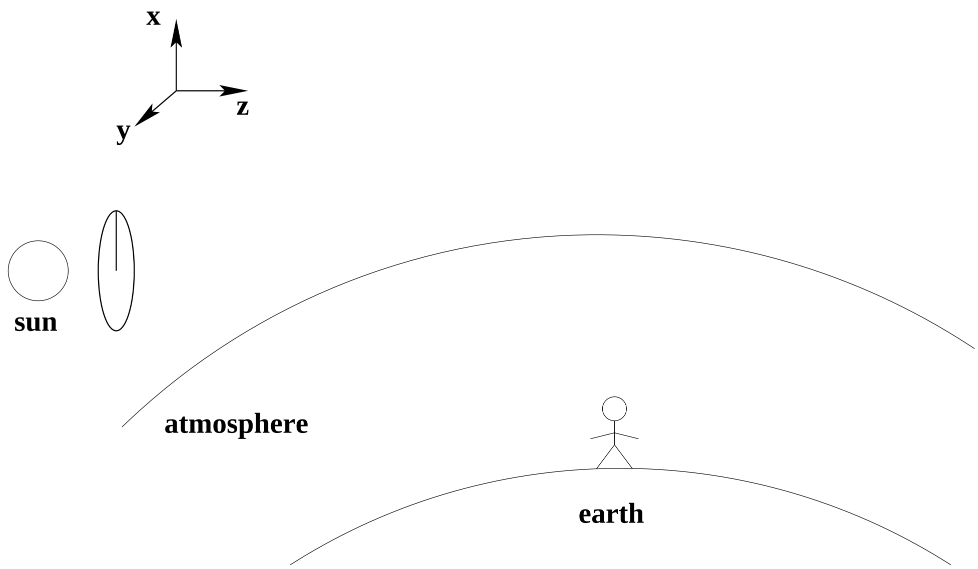


Figure 1: Dark above my head.

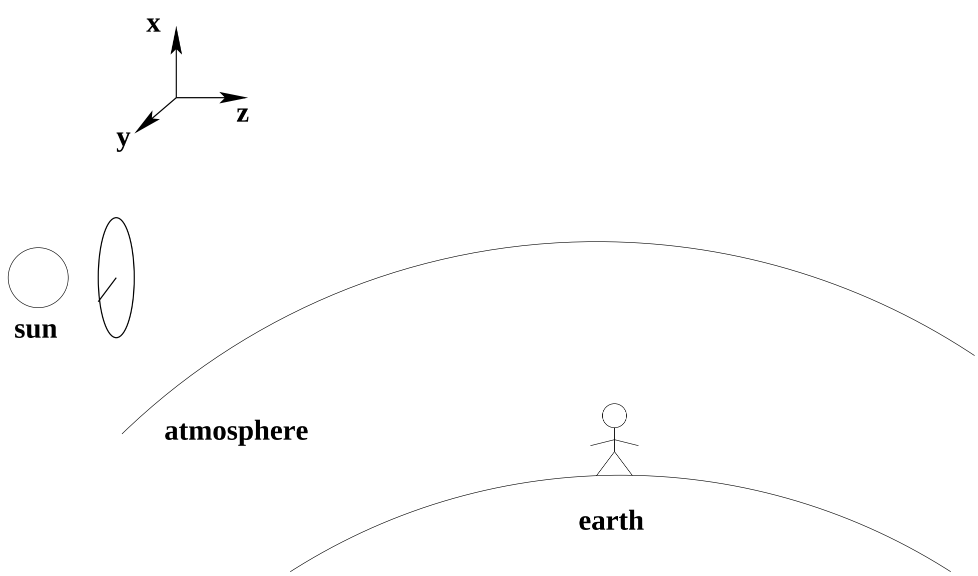


Figure 2: Bright above my head.

Something that has been missing so far is some discussion as to the *likelihood* that light will be scattered. The scattering process works roughly as follows: Light impinges on a molecule. Its electric field shakes that molecule's charges charges at the input frequency ω. The molecule then re-radiates this input light (often changing its direction, subject to the polarization constraints discussed above).

The electric field of the scattered light is determined by the *acceleration* of the charges in the scattering molecule — as we discussed back when we looked at special relativity, radiation comes from accelerating charges. This means that

$$E_{\text{scattered}} \propto \frac{\partial^2 d}{\partial t^2} \propto \omega^2$$

where d is the dipole moment of the shaken molecule. This moment will look something like $\cos \omega t$, which is how we get the ω^2 proportionality for the scattered electric field.

As we will discuss in much greater detail in our next lecture, the *intensity* or *brightness* of electromagnetic radiation is determined by the electric field *squared*:

$$I \propto E^2_{\text{scattered}} \propto \omega^4 \propto \lambda^{-4} \ .$$

The higher the radiation's frequency — or the smaller its wavelength — the higher the probability that it will be scattered.

Red light has a wavelength of about 700 nanometers. Blue light has a wavelength of about 350 nanometers. This means that blue light is scattered about $(700/350)^4 = 16$ times more effectively than red light!

This is why the sky is blue during the day: when we look away from the sun, the light we see is light that has scattered out of the initial propagation direction. The input light from the sun includes *all* wavelengths, but only the short wavelength stuff has a very high probability of scattering.

This is also why the sunset is red. At sunset, you are looking right at the sun — along the initial propagation direction — and you are looking through a large thickness of atmosphere. Most of the short wavelength stuff is already scattered out; only the long wavelength stuff remains.

21.5 Circular polarization

There is another polarization state that is encountered a lot in nature. Consider a wave that has the following form:

$$\begin{aligned}
\vec{E} &= E_0\hat{x}\sin(kz - \omega t) + E_0\hat{y}\cos(kz - \omega t) \\
\vec{B} &= E_0\hat{y}\sin(kz - \omega t) - E_0\hat{x}\cos(kz - \omega t) \ .
\end{aligned}$$

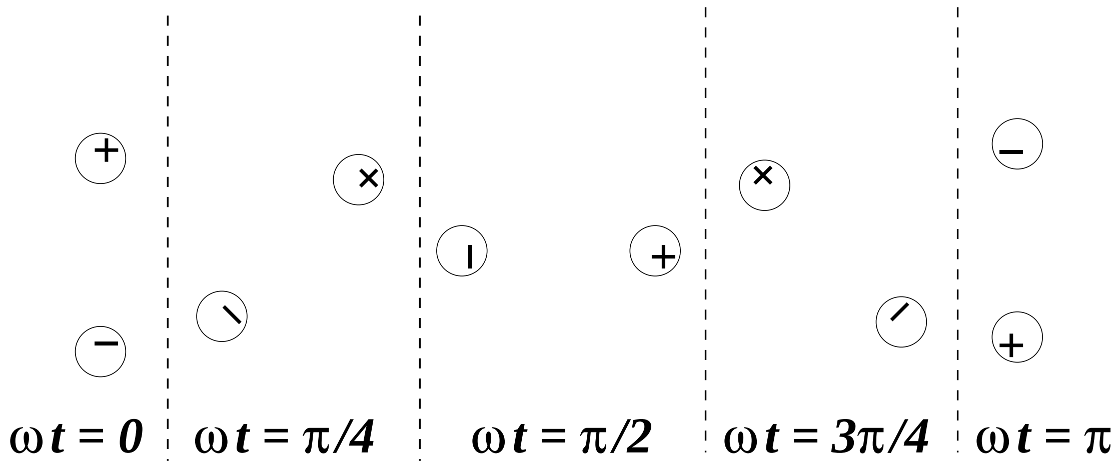
You should be able to show pretty easily that this satisfies all of our requirements for a plane wave. But what exactly is going on with this screwy form???

We can get some intuition for considering what this wave looks like at a fixed position. Suppose we focus on $z = 0$, so

$$\begin{aligned}
\vec{E} &= -E_0\hat{x}\sin(\omega t) + E_0\hat{y}\cos(\omega t) \\
\vec{B} &= -E_0\hat{y}\sin(\omega t) - E_0\hat{x}\cos(\omega t) \ .
\end{aligned}$$

The electric and magnetic fields *rotate* at the frequency ω! This is a state called *circular polarization*: the tip of the electric (and magnetic) field vector traces out a circle.

Circular polarization can often be particularly important because nature likes to produce it. Recall how we made linearly polarized radiation: we shook a charge configuration back and forth, with the charges oscillating along a simple line. It shouldn't surprise you to know that circular polarization is produced by a rotating charge configuration:



This rotating charge distribution would produce a circularly polarized wave that propagates out of the page.

The general polarization state is more usually somewhere between linear and circular polarization. Not surprisingly, this is called *elliptical* polarization. It is produced from rotating charge configurations that are viewed at some oblique angle. For example, if we looked at the configuration sketched above along its edge, we would get linearly polarized light. If we looked at it face on, we would get circular polarization. Elliptical polarization is what we get in the general case; it is intermediate to circular and linear polarization.

Finally, it's worth knowing that a linearly polarized state can be thought as a superposition of *two* circularly polarized states, but with their vectors propagating in opposite senses. Mathematically, this is so obvious that it may seem stupid. Let $\vec{E}_{\rm lin}$ denote a linearly polarized state; let $\vec{E}_{\rm CW}$ denote a state in which (at fixed location) the electric field is clockwise circular polarized; and let $\vec{E}_{\rm CCW}$ denote a state in which the electric field is counterclockwise circular polarized:

$$\begin{aligned}
\vec{E}_{\rm CW} &= -E_0\hat{x}\sin(kz-\omega t) + E_0\hat{y}\cos(kz-\omega t) \\
\vec{E}_{\rm CCW} &= E_0\hat{x}\sin(kz-\omega t) + E_0\hat{y}\cos(kz-\omega t) \\
\vec{E}_{\rm lin} &= \vec{E}_{\rm CW} + \vec{E}_{\rm CCW} = 2E_0\hat{y}\cos(kz-\omega t)\ .
\end{aligned}$$

This is so obvious you may wonder: why even bother belaboring the point? Well, the reason is that certain media interact with circularly polarized light in *very* interesting ways! For example, many organic molecules have a "chirality", meaning that they twist in a particular way. This turns out to mean that they interact with clockwise circular polarization *differently* than they interact with counterclockwise circular polarization!

If we were to shoot linearly polarized light into a medium containing such molecules, it would interact with the clockwise component of the linear wave differently than it would interact with the counterclockwise component. In other words, the medium would separate the linear polarization into its two circular components. We might expect any number of funky things to happen in this case.

Scott Hughes *6 May 2005*

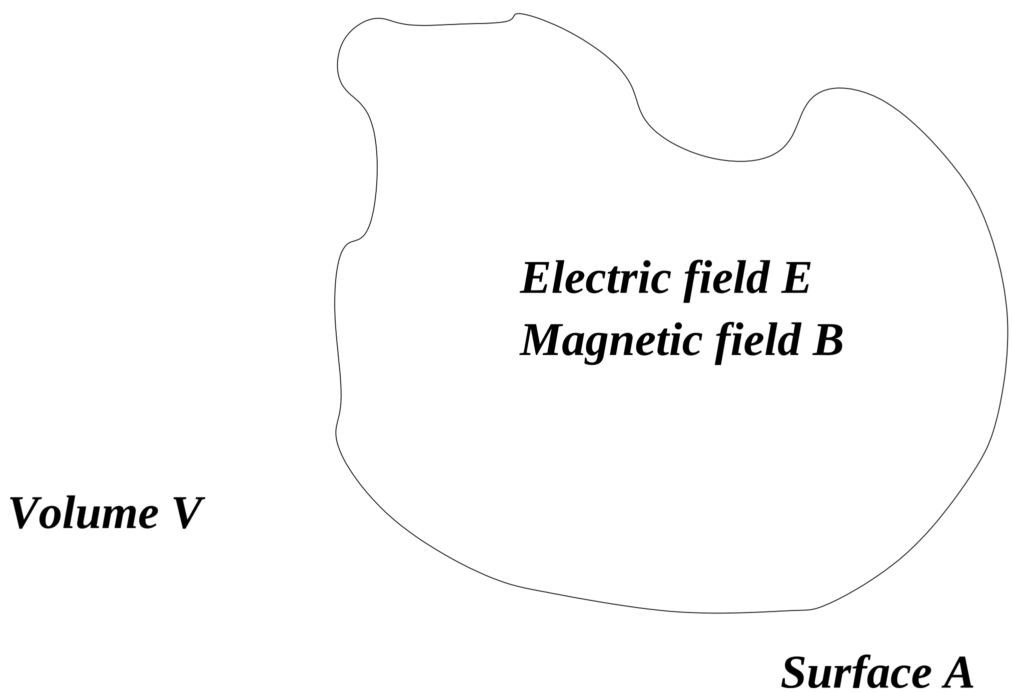
Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 22:
The Poynting vector: energy and momentum in radiation.
Transmission lines.

22.1 Electromagnetic energy

It is intuitively obvious that electromagnetic radiation carries energy — otherwise, the sun would do a pretty lousy job keeping the earth warm. In this lecture, we will work out how to describe the flow of energy carried by electromagnetic waves.

To begin, consider some volume V. Let its surface be the area A. This volume contains some mixture of electric and magnetic fields:



The electromagnetic energy density in this volume is given by

$$\frac{\text{energy}}{\text{volume}} = u = \frac{1}{8\pi}\left(\vec{E}\cdot\vec{E} + \vec{B}\cdot\vec{B}\right) \; .$$

We are interested in understanding how the total energy,

$$U = \int_V u\, dV = \frac{1}{8\pi}\int_V dV\left(\vec{E}\cdot\vec{E} + \vec{B}\cdot\vec{B}\right)$$

changes as a function of time. So, we take its derivative:

$$\frac{\partial U}{\partial t} = \frac{\partial}{\partial t}\frac{1}{8\pi}\int_V dV\left(\vec{E}\cdot\vec{E} + \vec{B}\cdot\vec{B}\right) \; .$$

Since we have fixed the integration region, we can take the $\partial/\partial t$ ender the integral:

$$\begin{aligned}\frac{\partial U}{\partial t} &= \frac{1}{8\pi}\int_V dV \frac{\partial}{\partial t}\left(\vec{E}\cdot\vec{E}+\vec{B}\cdot\vec{B}\right)\\ &= \frac{1}{4\pi}\int_V dV\left(\vec{E}\cdot\frac{\partial\vec{E}}{\partial t}+\vec{B}\cdot\frac{\partial\vec{B}}{\partial t}\right)\ .\end{aligned}$$

Rearranging the source free Maxwell equations,

$$\frac{\partial\vec{E}}{\partial t} = c\vec{\nabla}\times\vec{B}$$
$$\frac{\partial\vec{B}}{\partial t} = -c\vec{\nabla}\times\vec{E}$$

we can get rid of the $\vec{E}$ and $\vec{B}$ time derivatives:

$$\frac{\partial U}{\partial t} = \frac{c}{4\pi}\int_V dV\left[\vec{E}\cdot\left(\vec{\nabla}\times\vec{B}\right)-\vec{B}\cdot\left(\vec{\nabla}\times\vec{E}\right)\right]\ .$$

22.2 The Poynting vector

The expression for $\partial U/\partial t$ given above is as far as we can go without invoking a vector identity. With a little effort, you should be able to prove that

$$\vec{\nabla}\cdot\left(\vec{E}\times\vec{B}\right) = -\vec{E}\cdot\left(\vec{\nabla}\times\vec{B}\right)+\vec{B}\cdot\left(\vec{\nabla}\times\vec{E}\right)\ .$$

Comparing with our expression for $\partial U/\partial t$, we see that this expression simplifies things:

$$\begin{aligned}\frac{\partial U}{\partial t} &= -\frac{c}{4\pi}\int_V dV\,\vec{\nabla}\cdot\left(\vec{E}\times\vec{B}\right)\\ &\equiv -\int_V dV\,\vec{\nabla}\cdot\vec{S}\ .\end{aligned}$$

On the second line we have defined

$$\vec{S} = \frac{c}{4\pi}\vec{E}\times\vec{B}\ ;$$

we'll discuss this vector in greater detail very soon.

Since we have a volume integral of a divergence, the obvious thing to do here is to apply Gauss's theorem. This changes the integral over V into an integral over the surface A:

$$\frac{\partial U}{\partial t} = -\int_A d\vec{a}\cdot\vec{S}\ .$$

In other words, **The rate of change of the electromagnetic energy in the volume V is given by the flux of $\vec{S}$ through V's surface.** Notice the minus sign; this is easily explained. Recall that $d\vec{a}$ points outward. If $\vec{S}$ and $d\vec{a}$ are parallel, energy is *leaving* the system, so U decreases. And vice versa.

$\vec{S}$ is called the Poynting[1] vector. It points[2] in the same direction as the electromagnetic energy flow. (This gives a somewhat more physical picture of the fact that $\vec{E}\times\vec{B}$ tells us the

[1]It's actually named after a person, John Henry Poynting. You've got to wonder if he was fated to work this quantity out.

[2]This is usually when a lesser lecturer would make a terrible pun about the direction in which the flow of energy "Poynts". I'll spare you.

propagation direction for electromagnetic radiation.) As noted above, the flux of $\vec{S}$ through some surface A gives the total *power* flowing through A:

$$\text{Power through } A = \int_A \vec{S} \cdot d\vec{a} \; .$$

22.2.1 Dimensional analysis

The units of the Poynting vector are

$$\begin{aligned} \text{Poynting} &\equiv \text{velocity} \times \text{electric field} \times \text{magnetic field} \\ &\equiv \text{velocity} \times \text{electric field}^2 \end{aligned}$$

Electric field squared gives us energy per unit volume, so

$$\begin{aligned} \text{Poynting} &\equiv \frac{\text{length}}{\text{time}} \times \frac{\text{energy}}{\text{length}^3} \\ &= \frac{\text{power}}{\text{area}} \; . \end{aligned}$$

This is just what we should have if the flux of $\vec{S}$ is to be the power through an area. In cgs units, $\vec{S}$ has the units of erg/(s-cm^2).

It's worth noting at this point that the magnitude of $\vec{S}$ is a quantity known as the intensity, I. A very intense source of radiation is something that emits lots of power into a very small area.

22.3 Examples

22.3.1 Plane wave

Let's look at a linearly polarized plane wave, $\vec{E} = E_0\hat{x}\sin(kz - \omega t)$, $\vec{B} = E_0\hat{y}\sin(kz - \omega t)$. The Poynting vector associated with this wave is

$$\vec{S} = \frac{c}{4\pi}\vec{E} \times \vec{B} = \frac{c}{4\pi}E_0^2\sin^2(kz - \omega t)\hat{k} \; .$$

It's useful to compare this to the energy density in the wave:

$$\begin{aligned} u &= \frac{1}{8\pi}\left(\vec{E} \cdot \vec{E} + \vec{B} \cdot \vec{B}\right) \\ &= \frac{1}{4\pi}E_0^2\sin^2(kz - \omega t) \end{aligned}$$

Comparing, we see that

$$\vec{S} = u\vec{c}$$

where $\vec{c} = c\hat{k}$ is the vectorial velocity of the radiation. This is very nicely in keeping with the idea that $\vec{S}$ tells us about the flow of energy! (Note the similarity to the relation between current density and the velocity of moving charges, $\vec{J} = \rho\vec{v}$. $\vec{S}$ is a flow of energy density in the same way that $\vec{J}$ is a flow of charge density.)

For many kinds of radiation that we will wish to study, k and ω are so large that $\sin^2(kz - \omega t)$ oscillates extremely quickly. (For example, the frequency of visible light —

mean wavelength roughly 500 nanometers — is about 6×10^{14} Hz.) It makes a lot more sense to take the average of this quantity, which amounts to replacing the $\sin^2$ with 1/2:

$$\langle \vec{S} \rangle = \frac{c\hat{k}}{8\pi} E_0^2 \,.$$

The intensity (magnitude of $\vec{S}$) is most typically discussed in terms of the averaged Poynting vector:

$$I = |\langle \vec{S} \rangle| = \frac{c}{8\pi} E_0^2 \,.$$

22.3.2 Spherical wavefronts

If we solve the wave equation in spherical coordinates, one important solution is given by

$$\begin{aligned} \vec{E} &= \frac{\omega^2 p}{c^2} \sin\theta \frac{\sin(kr - \omega t)}{r} \hat{\theta} \\ \vec{B} &= \frac{\omega^2 p}{c^2} \sin\theta \frac{\sin(kr - \omega t)}{r} \hat{\phi} \,. \end{aligned}$$

(This solution is actually only good for distances $r \gg \lambda$, where $\lambda = 2\pi/k$ is the usual wavelength. This is a subtlety far beyond the scope of 8.022! For now, just trust me.)

This radiation solution is what we find for an oscillating dipole arranged so that the dipole vector, $\vec{p}$, lies parallel to the z axis (p is the magnitude of $\vec{p}$). Notice that the radiation propagates out *radially*: although there is some angular dependence, the radiation flows out through spheres.

The Poynting vector we construct for this wave is

$$\begin{aligned} \vec{S} &= \frac{c}{4\pi} \vec{E} \times \vec{B} \\ &= \frac{1}{4\pi c^3} \omega^4 p^2 \sin^2\theta \frac{\sin^2(kr - \omega t)}{r^2} \hat{r} \,. \end{aligned}$$

Averaging over time, we have

$$\langle \vec{S} \rangle = \frac{\omega^4 p^2}{8\pi c^3 r^2} \sin^2\theta \hat{r} \,.$$

The Poynting vector (and hence the intensity) fall of with a $1/r^2$ law. This is extremely important! Suppose we draw a sphere of radius R around this system, centered on the origin (where the oscillating dipole is located). Let's now compute the total rate of energy flowing through this sphere:

$$\left\langle \frac{\partial U}{\partial t} \right\rangle = \int_R d\vec{a} \cdot \langle \vec{S} \rangle \,.$$

The area element on this sphere is $d\vec{a} = R^2 \sin\theta \, d\theta \, d\phi \, \hat{r}$; the integral becomes

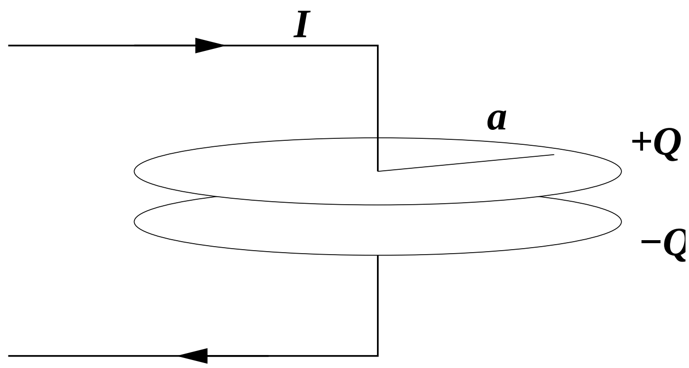
$$\begin{aligned} \left\langle \frac{\partial U}{\partial t} \right\rangle &= \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \left(R^2 \sin\theta \hat{r}\right) \cdot \left(\frac{\omega^4 p^2}{8\pi c^3 R^2} \sin^2\theta \hat{r}\right) \\ &= \frac{\omega^4 p^2}{8\pi c^3} \int_0^{2\pi} d\phi \int_0^{\pi} \sin^3\theta \, d\theta \\ &= \frac{\omega^4 p^2}{4c^3} \int_0^{\pi} \sin^3\theta \, d\theta \\ &= \frac{\omega^4 p^2}{3c^3} \,. \end{aligned}$$

This final result is a special case of what is known as the *Larmor formula*. Notice that this result is *completely independent of the radius* R! What happened is that the Poynting vector or intensity drops off like $1/r^2$. However, the total area of a spherical surface grows with r^2, so the *total rate of energy flow* is constant.

This, in essence, is telling us that electromagnetic waves conserve energy. They are spreading out over large areas, but the total energy they carry remains constant.

22.3.3 Capacitor

The Poynting vector applies in *any* situation in which both an electric and a magnetic field appear, not just with radiation. In this example, we will apply it to the charging up of a capacitor. Consider a capacitor that currently has a charge separation of Q, and has a current I flowing onto it:



The electric field between the plates is

$$E = 4\pi Q/A = 4Q/a^2 \ ;$$

it points down. From the generalized Ampere's law, we have

$$\oint \vec{B} \cdot d\vec{s} = \frac{1}{c}\frac{d\Phi_E}{dt}$$

We take the contour integral around a ring of radius r; the flux is computed through this ring:

$$\begin{aligned} 2\pi r B(r) &= \frac{\pi r^2}{c}\frac{dE}{dt} \\ &= \frac{\pi r^2}{c}\frac{4}{a^2}\frac{dQ}{dt} \\ &= \frac{4\pi I r^2}{ca^2} \ . \end{aligned}$$

So the magnetic field at radius r from the center of the capacitor is

$$B(r) = \frac{2Ir}{ca^2} \ .$$

The direction is of course circulational, defined by right hand rule with your thumb pointing down. This means that at the front of the capacitor $\vec{B}$ points to the left; at the back, it points to the right; etc.

The cool thing is to now evaluate the Poynting vector: since $\vec{E}$ and $\vec{B}$ are orthogonal, we clearly have

$$S(r) = \frac{c}{4\pi} EB = \frac{2QI}{\pi a^4} \ .$$

What is the direction? At the front, we have (down)×(left). This points *into* the capacitor. At the back, we have (down)×(right). This again points *into* the capacitor. In fact, we can quickly show that at *all* points, $\vec{S}$ points into the capacitor.

This makes perfect sense! It *should* point inwards since, as we well know, the amount of energy in the capacitor increases as the plates charge up. The Poynting vector provides us with a way of visualizing this.

You will explore this facet of the Poynting vector further on Pset 11.

22.4 Momentum carried by radiation

Since electromagnetic radiation carries energy, it won't surprise you to learn that it carries momentum as well. A simple way to see this is to make use of the relativistic mass-energy-momentum formula you worked out in Pset 6:

$$E^2 = |\vec{p}|^2 c^2 + m^2 c^4 \ .$$

If we plug in $m = 0$, we find

$$E = |\vec{p}| c \ ; \qquad |\vec{p}| = E/c \ .$$

Let's think about using the relationship with what we've worked out so far. The Poynting vector tells us about the rate at energy is delivered to a unit area, just as we learned in the dimensional analysis discussion above. If we divide by c, then what we must have is the rate at which *momentum* is delivered to a unit area:

$$\frac{\vec{S}}{c} = \frac{\vec{E} \times \vec{B}}{4\pi}$$

Dimensionally,

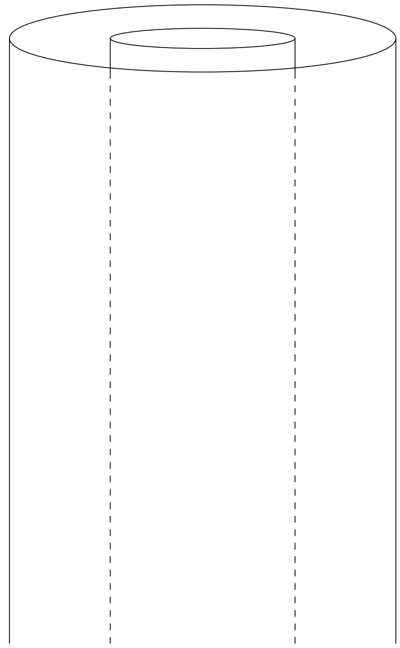
$$\frac{\vec{S}}{c} = \frac{\text{momentum}}{\text{time} \times \text{area}} = \frac{\text{force}}{\text{area}} = \text{pressure} \ .$$

Radiation exerts pressure! Intensity divided by the speed of light tells us how much pressure is exerted by some source of light.

22.5 Transmission lines

Suppose we want to send a signal from one device to another — our goal is to get electromagnetic energy from point A to point B. What's the best way to do this? If we just use wires, we're going to have problems — the wires will act like antennas and we'll lose a lot of energy to radiation. Instead, we would like to set up some kind of structure that *shields* the signal, so that no field leaks outside. This setup is called a transmission line.

The simplest example of a transmission line is a coaxial cable: a pair of conducting tubes nested in one another. Equal and opposite currents flow on the inner and the outer tubes,



so that the net current "seen" outside the cylinder is zero. Note that this is *not* a simple DC current — it is a highly oscillatory AC current. This means that there must be some charge distribution as well. Equal and opposite charges "live" on the inner and outer tubes of the coax.

Suppose that the inductance per unit length of this cable is L', and that its capacitance per unit length is C'. Consider a short length Δx of this cable. This little segment has a total inductance of $L'\Delta x$. If the current flowing down the cable is I and has the rate of change $\partial I/\partial t$, then the voltage drop due to inductance over this short length is

$$\Delta V = V(x + \Delta x) - V(x) = -(L'\Delta x)\frac{\partial I}{\partial t} .$$

Divide both sides by Δx and take the limit:

$$\frac{\partial V}{\partial x} = -L'\frac{\partial I}{\partial t} .$$

Consider now the charge on this little segment: since it is at potential $V(x)$, and its capacitance is $C'\Delta x$, the amount of charge on this segment is

$$\Delta Q = (C'\Delta x)V(x) .$$

Divide and take the limit:

$$\frac{\partial Q}{\partial x} = C'V(x) .$$

Let's take another derivative of this:

$$\begin{aligned}
\frac{\partial^2 Q}{\partial x^2} &= C'\frac{\partial V}{\partial x} \\
&= -C'L'\frac{\partial I}{\partial t} \\
&= C'L'\frac{\partial^2 Q}{\partial t^2} .
\end{aligned}$$

In going from the first to the second line, we used our previous result relating $\partial V/\partial x$ and $\partial I/\partial t$; in going from the second to the third line, we used $I = -\partial Q/\partial t$ (discharging capacitor current).

Rearranging this, we see yet another wave equation!

$$\frac{\partial^2 Q}{\partial t^2} - \frac{1}{C'L'}\frac{\partial^2 Q}{\partial x^2} = 0\ .$$

With a little bit of tweaking, you should be able to convince yourself that this equation holds replacing Q with V and I as well. The speed of propagation of this wave is

$$v = \frac{1}{\sqrt{L'C'}}\ .$$

Now, for the coaxial cable introduced above, you should be able to show very easily that

$$\begin{aligned} C' &= \frac{1}{2\ln(b/a)} \\ L' &= \frac{2\ln(b/a)}{c^2}\ . \end{aligned}$$

(If you can't work it out, go back to old psets! You worked out the capacitance in Pset 4 and the inductance in Pset 8; you just to divide out the length l to put things on a "per unit length" basis.) This means that

$$v = \frac{1}{\sqrt{L'C'}} = c\ .$$

The wave propagates down the transmission line at the speed of light! Not surprising, at least in retrospect.

One other very important quantity determines the characteristic of a transmission line: its impedance. Following our intuition from our study of AC circuits, we define the impedance as the ratio of the voltage to the current:

$$Z = \frac{V}{I} = \frac{V}{\Delta Q/\Delta t}\ .$$

As we already discussed, for a small segment of the cable $\Delta Q = (C'\Delta x)V$, so

$$\frac{\Delta Q}{\Delta t} = C'V\frac{\Delta x}{\Delta t} = \frac{C'V}{\sqrt{L'C'}} = V\sqrt{\frac{C'}{L'}}$$

where we have used the fact that $\Delta x/\Delta t$ is just the speed of propagation down the line.

Putting everything together, we find

$$Z = \sqrt{\frac{L'}{C'}}\ .$$

For the coaxial cable, this yields

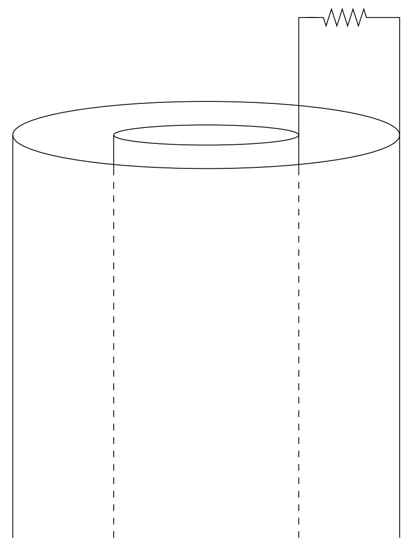
$$Z = \frac{2\ln(b/a)}{c}\ .$$

Using the conversion factor 1 Ohm $= 1.1 \times 10^{-12}$ sec/cm, this means

$$Z = 60\,\text{Ohms}\,\ln(b/a)\ .$$

Many coaxial cables are actually filled with a dielectric material (which we have not discussed in detail); this has the effect of increasing the capacitance, which reduces the line's characteristic impedance. 50 Ohms is a value that is commonly encountered (e.g., in cable TV systems).

One of the most important consequences of the impedance of a transmission line is in determining how it must be *terminated.* Roughly speaking, a terminated cable has the form



The resistor might actually be some electronic device, like a television or a computer; then again, it might just be a resistor. The point is that the signal which is flowing down the cable is then fed into this device or resistor.

If the resistance of this "load" precisely matches the impedance of the cable, then we have *matched* the impedances. Because V/I does not change in going from the cable to the resistor, it is as though the cable were of infinite length! There is no way for the signal to "know" whether it is in the resistor or in the cable.

Suppose the resistance were ∞ Ohms — an open circuit. Then, the currents would essentially bounce off the "resistor" and head back where they came from. The signal would just reverse course — you'd get a horrible reflection back up the line. Suppose the resistance were 0 Ohms — a short circuit. Then, the currents would just flow from the inner tube to the outer tube, and vice versa — the signal would reverse course *and* switch sign on its amplitude!

In general, if the impedance of the cable doesn't match the impedance of the load, you get reflections of this sort. This is one reason why cheap cable TV equipment can give you a horrible picture — there are these extra signals bouncing around, which eventually arrive at the TV late, and possible with a 180° phase shift. This causes ghost images to appear, and generally confuses your equipment.

Scott Hughes *10 May 2005*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF PHYSICS
8.022 SPRING 2005

LECTURE 23:
MAGNETIC MATERIALS

23.1 Magnetic fields and stuff

At this stage of 8.022, we have essentially covered *all* of the material that is typically taught in this course. Congratulations — you now have a solid grounding in all the major concepts of electricity and magnetism!

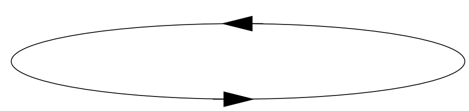
There remains one major, important subject which we have not discussed in depth: the interaction of magnetic fields and materials. At this point, we don't even have a really good understanding of a bar magnet works — an unsatisfying state of affairs, given that this is the way in which we normally encounter magnetism in the real world!

One reason we have avoided covering this subject is that it is not really possible to discuss it properly within getting into a detailed discussion of the quantum mechanical description of matter. The way in which matter responds to magnetic fields is *totally* determined by the quantum mechanical nature of their molecular structure, particularly their electrons.

Nonetheless, we can make significant headway in understanding the interaction of magnetic fields and materials by combining what we have learned so far with a somewhat approximate, qualitative description of how materials respond to magnetic fields. The main concepts we will need are summarized in the following two subsections:

23.1.1 Electron orbitals

The electrons in a molecule exist in *orbits*. *Very roughly*, we can picture an orbit as a simple loop of current:



A loop of current like this of course tends to generate its own magnetic field. In most materials, there are an enormous number of orbits like this, randomly oriented so that they produce no net field. Suppose an external magnetic field is applied to some substance that contains many orbitals like this. The net tendency can be understood in terms of Lenz's law: the orbits "rearrange" themselves in order to *oppose* the change in magnetic flux. This tendency for Lenz's law to work on the microscopic scale ends up *opposing the magnetic field from the material.*

23.1.2 Intrinsic magnetic moment of the electron

One other quantum mechanical property of electrons plays an extremely important role in this discussion: electrons have a built-in, *intrinsic* magnetic moment. Roughly speaking, this means that each electron *all on its own* acts as a source of magnetic field, producing a dipole-type field very similar to that of current loop.

Because this field is associated with the electron itself, it does not exhibit the Lenz's law type behavior of the field that we see from the orbits. Instead, the most important behavior in this context is the fact that a magnetic moment $\vec{m}$ placed in an external field $\vec{B}$ feels a *torque*:

$$\vec{N} = \vec{m} \times \vec{B} \,.$$

In this case, the action of this torque tends to line up the electrons' magnetic moments with the external field. In this case, you find that *the magnetic field is augmented within the material.*

23.1.3 Which wins?

To summarize the above discussion, the quantum mechanical nature of electrons in molecules leads to two behaviors:

- Lenz's law on the scale of electron orbitals *opposes* magnetic fields from entering a material.

- Magnetic torque acting on the individual electrons *augments* magnetic fields in a material.

These two behaviors are in complete opposition to one another! However, both occur, and both are important. A natural question to ask is: *Which* of these two behaviors is *more* important?

The answer to this question varies from material to material, depending upon its detailed electronic orbital structure. The first property — Lenz's law on the orbital scale — plays some role in *all* materials. In many cases, this is the end of the story. Such materials are called *diamagnetic*. A diamagnetic material is one whose *magnetization* (to be defined precisely in a moment) opposes an external magnetic field. Even though almost all materials are diamagnetic, the effect is so puny that it is typically very difficult to see. A consequence of this is that a diagmagnetic substance will be expelled from a magnetic field.

For some materials, the second property — alignment of electron magnetic moments — wins out. Such materials typically have several electron orbits that contain unpaired electrons; the orbit thus has a net magnetic moment. (In most diamagnetic materials, the orbits are filled with paired electrons, so that the orbit has no net moment[1].) Such materials are called *paramagnetic*. A paramagnetic material has a magnetization that augments an external field. Paramagnetic substances are pulled into a magnetic field. In the vast majority of cases, paramagnetic effects are also so puny that they can barely be seen[2].

One particular class of materials acts essentially as paramagnetic materials do. However, they respond *so* strongly that they truly belong in a class of their own. These are the *ferromagnetic* materials.

[1] Bear in mind that this is a *highly* qualitative discussion!

[2] There are some interesting exceptions. Oxygen, O_2, turns out to be paramagnetic. However, the random motion of gaseous oxygen wipes out any observable effect. If we cool it to a liquid state, though, the paramagnetism is easily seen.

23.2 Some definitions

A couple of definitions are needed to describe what we are going to discuss. Foremost is the *magnetization* $\vec{M}$: the magnetic dipole moment per unit volume of a material. Notice that magnetization has the same dimensions as magnetic field:

$$\begin{aligned}
\vec{M} &= \frac{\text{magnetic moment}}{\text{volume}} \\
&= \frac{\text{current } \times \text{ area/velocity}}{\text{length}^3} \\
&= \frac{\text{current/velocity}}{\text{length}} \,.
\end{aligned}$$

A material with constant magnetization $\vec{M}$ and volume V acts just like a magnetic dipole with $\vec{m} = \vec{M}V$.

The second important definition is a bit of a wierd one: it is a new kind of magnetic field denoted $\vec{H}$. This $\vec{H}$ is defined in terms of the "normal" magnetic field $\vec{B}$ and the magnetization $\vec{M}$ via

$$\vec{B} = \vec{H} + 4\pi\vec{M} \,.$$

In other words, when there is no magnetization, the "normal" magnetic field $\vec{B}$ is just this new field $\vec{H}$. This means that $\vec{H}$ is nothing more than the bit of the magnetic field that arises from "normal" electric currents, as opposed to those that are "bound" in matter.

23.2.1 More details on $\vec{H}$

We can do some more stuff with $\vec{H}$, though this is really beyond the scope of what we need for our discussion. If we take the curl of $\vec{H}$, we get

$$\vec{\nabla} \times \vec{H} = \frac{4\pi}{c}\vec{J}_{\rm free}$$

where $J_{\rm free}$ is the density of *free* electrical current. The curl of the magnetization defines *bound* currents:

$$\vec{\nabla} \times \vec{M} = \vec{J}_{\rm bound}/c \,.$$

Putting these together, we have

$$\begin{aligned}
\vec{\nabla} \times \vec{B} &= \vec{\nabla} \times \vec{H} + 4\pi\vec{\nabla} \times \vec{M} \\
&= \frac{4\pi}{c}\vec{J}_{\rm free} + \frac{4\pi}{c}\vec{J}_{\rm bound} \\
&= \frac{4\pi}{c}\vec{J} \,.
\end{aligned}$$

In other words, the *total* current density is given by adding the *free* current density (the current due to moving charges that we can imagine controlling) and the *bound* current density (the current that is "bound up" in electron orbits, and is really a part of the material).

23.3 Magnetic susceptibility

A wide variety of materials exhibit *linear magnetization*, meaning that the magnetization $\vec{M}$ depends linearly on the field $\vec{H}$:

$$\vec{M} = \chi_m \vec{H} \ .$$

The quantity χ_m is the *magnetic susceptibility*. From this, we have

$$\vec{B} = \vec{H}\left(1 + 4\pi\chi_m\right) \ .$$

If $\chi_m < 0$, the magnetic field is decreased by the material. This corresponds to *diamagnetism*. If $\chi_m > 0$, the magnetic field is increased by the material. This corresponds to *paramagnetism*.

23.4 Ferromagnetism

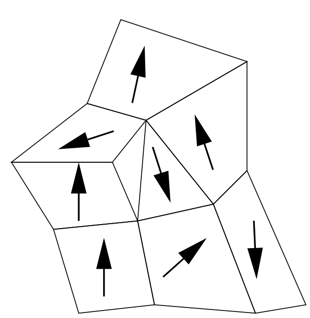
Roughly speaking, ferromagnetism is paramagnetism on steroids. Its key feature is that it is *nonlinear*: the $\vec{M}$ and $\vec{H}$ do *not* have a simple linear relation between one another. This leads to some surprising results. In particular, it means that once the "external" field $\vec{H}$ is turned off, the magnetization $\vec{M}$ remains! This of course is how permanent magnets are made.

The reasons for this nonlinear behavior are surprisingly complicated, and are difficult to explain correctly in a simple manner. (It is ironic that permanent magnets, probably the most familiar manifestation of magnetism, turn out to be so difficult to explain!) It isn't so hard to describe the most important facets of ferromagnetism qualitatively; the hard part is explaining *why* those qualitative behaviors emerge. That, sadly, is far beyond the scope of 8.022.

23.4.1 Ferromagnetic domains

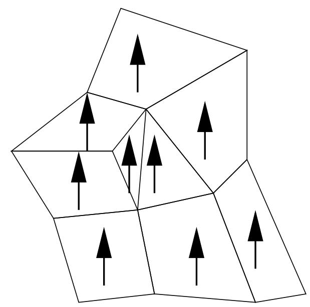
Since ferromagnetism is conceptually a lot like paramagnetism, we expect that it will be explainable in terms of the alignment of many atomic scale magnetic moments. This is indeed the case. However, with ferromagnetism there is an important modification: one finds that the moments of *many* atoms or molecules tend to be aligned in small regions. In "normal" paramagnetic materials, the moments are randomly arranged until an external field comes along and aligns them.

If you were to look at the magnetic moment layout of a typical ferromagnetic materials, you would find something like this:



Enough atomic scale moments are aligned that definite regions or *domains* of magnetization exist. Each domain is fairly small, and each is randomly oriented with respect to its neighbor; the material as a whole has no net magnetization. This is why a random piece of iron (a representative ferromagnetic material) is usually not magnetized.

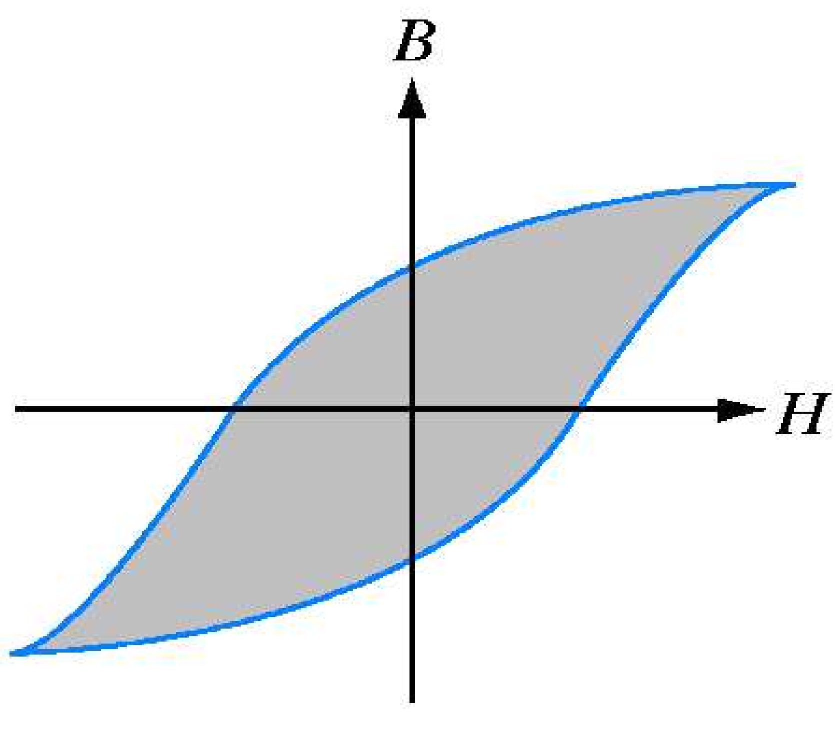
If this material is placed into a strong enough external magnetic field, the domains are re-aligned so that they all point in the same direction:



Once the external magnetic field is removed, it turns out to be energetically favorable for these domains to remain pointing in this direction. The domains are fixed with this magnetization. This is how permanent magnets are made.

23.4.2 Hysteresis

The nonlinear relationship between $\vec{H}$ and $\vec{B}$ means that the magnetic field $\vec{B}$ in a ferromagnetic material has a rather complicated dependence on the "external" applied field $\vec{H}$. As $\vec{H}$ is varied between a minimum and a maximum, one finds that $\vec{B}$ in the material traces out an interesting shape called a *hysteresis curve*:



Notice that when $H = 0$, $B \neq 0$ — reflecting the fact that the material has been given a permanent magnetization. You may wonder looking at this plot which of the two possible values it will take — the positive B or the negative one? The answer depends on the history of the magnetization. If H was large and positive most recently, B will go to the positive value when H is set to zero. If H was large and negative, B will go to the negative value.

It turns out (though it's a bit beyond the scope of this course to show this) that it takes work to "go around" the hysteresis curve. As we go from $+H$ to $-H$, we have to flip lots of magnetic domains around — it takes energy to do that! The amount of work that is done in going from $+H \to -H \to +H$ is proportional to the area of the hysteresis curve. One consequence of this is that solenoids with an iron core can be very *lossy* elements in AC circuits. An AC circuit is continually changing the direction of flow of the current, and hence of the $\vec{H}$ field. If a portion of the circuit contains a ferromagnet, some of the power put into the circuit will be lost by continually changing the orientation of the ferromagnet's domains.

23.4.3 Curie temperature

No discussion of ferromagnetism is complete without mentioning the *Curie temperature*. The Curie temperature, T_C, is a temperature above which a material *very suddenly* ceases to act like a ferromagnetic. If the temperature is above T_C, the random motion of the magnetic moments is so strong that they can *never* become aligned. Just a tiny bit below, though, and they can align! It's somewhat amazing how strong the transition is at this temperature. For iron, $T_C = 770^\circ$ Celsius.

Scott Hughes *12 May 2005*

Massachusetts Institute of Technology
Department of Physics
8.022 Spring 2005

Lecture 24:
A (very) brief introduction to general relativity.

24.1 Gravity?

The Coulomb interaction between two point charges looks essentially identical to the gravitational interaction between two masses. Does this mean that everything we have done so far can also be used to describe gravitational interactions (perhaps with some slight modifications)?

The answer to this turns out to be *almost* — but not quite. To understand what this means in detail, we need to do a little bit of background work. The key point to be made here is that Maxwell's equations and electricity and magnetism are relativistically correct: all of E&M "knows about" special relativity, in the sense that the equations make sense in any frame of reference. (When we transform reference frames, we modify the $\vec{E}$ and $\vec{B}$ fields according to those relativistic transformation laws we all learned to love ages ago. However, the new fields *still* satisfy Maxwell equations.) This is *NOT* the case for gravity! Significant modifications need to be made to make gravity relativistic. What we end up with is Einstein's theory of *general relativity*. In this lecture we will look at some basic properties of general relativity.

To begin, we need to introduce a bit of notation.

24.2 4-vectors

Since space and time are unified in relativity, it doesn't make much sense to treat them separately. Accordingly, many concepts are described using *4-vectors*, quantities that are essentially just like the vectors we've known and loved since kindergarten, but with an extra "timelike" component. For example, the position 4-vector is written

$$x^\mu = (ct, x, y, z)\ .$$

The way to read this is that μ is an *index*, ranging from 0 to 3: $x^0 = x^{\mu=0} = ct$; $x^1 = x^{\mu=1} = x$; $x^2 = x^{\mu=2} = y$; $x^3 = x^{\mu=3} = z$. Whenever you see an x with a number superscript, it means the index — it does NOT mean take it to a higher power! This system can be a little confusing when you first encounter it.

That the "timelike component of position" is time makes perfect sense. What do we get when we try to make something like momentum into a 4-vector? It turns out that the "timelike component of momentum" is the *energy*:

$$p^\mu = (E/c, p^x, p^y, p^z)\ .$$

We'll introduce a few other 4-vectors before we're done here.

24.2.1 Invariant interval

Back on Pset 6, we showed that the quantity

$$\Delta s^2 = -c^2\Delta t + \Delta x^2 + \Delta y^2 + \Delta z^2$$

is *invariant*: **all** inertial observers agree on its value. This can be written this in terms of a quantity called the *metric*:

$$\Delta s^2 = \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} g_{\mu\nu}\Delta x^\mu \Delta x^\nu$$

where the *metric tensor* $g_{\mu\nu}$ can be represented as a 4×4 matrix:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

This relation is usually written using the *Einstein summation convention*: repeated indices are *assumed* to be summed from 0 to 3. In this convention, the invariant interval then becomes

$$\Delta s^2 = g_{\mu\nu}\Delta x^\mu \Delta x^\nu .$$

We will use this notation from now on.

24.2.2 Covariant versus contravariant

The 4-vectors that we have introduced so far are, technically speaking, written in terms of *contravariant* components. Using the metric, we define a new version of our 4-vector:

$$\begin{aligned} x_\mu &\equiv g_{\mu\nu}x^\nu \\ &= (-ct, x, y, z) . \end{aligned}$$

The effect is just to change the sign of the timelike component.

It should be noted that there is *no* physics in this operation. The difference between "index up" (contravariant) and "index down" (covariant) is just needed to make sure that the math works out right. In particular, in writing down quantities that matter for physics, you typically end up summing over repeated indices. When we do this, the index has to be down on one quantity and up on the other.

24.2.3 Derivatives

Finally, one handy notational definition: we write

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} , \qquad \partial^\mu \equiv \frac{\partial}{\partial x_\mu} .$$

There's absolutely nothing deep about this! However, it makes certain quantities look nice.

Notice that $\partial_0 = (1/c)\partial/\partial t = -\partial^0$.

24.3 Electric & magnetic fields?

Can we make a 4-vector that incorporates electric and magnetic fields? With a little thought, we see that the answer to this has to be **no** — the three components of electric fields and the three components of magnetic fields are 6 separate numbers. A 4-vector — pretty much by definition! — only has 4 separate numbers.

What we *can* do is make a 4-vector out of the magnetic vector potential. We use the electrostatic scalar potential as the time component, and find

$$A^\mu = (\phi, A^x, A^y, A^z) .$$

(Note that in cgs, ϕ and $\vec{A}$ have the same units.) How do we get fields from this potential? We take a derivative! The one that we want turns out to be

$$F^{\mu\nu} \equiv \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu} .$$

This tensor contains EXACTLY the quantities that we want! When we write out the tensor $F_{\mu\nu}$, we see that its components are given by

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix} .$$

The tensor is often called the *Faraday tensor*.

Although it's not so simple to see where this one comes from, there's a second tensor we can construct in a similar manner, often just called the *Dual Faraday* tensor:

$$^*F^{\mu\nu} = \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z & E_y \\ -B_y & E_z & 0 & -E_x \\ -B_z & -E_y & E_x & 0 \end{pmatrix}$$

We'll discuss the point of introducing this "dual" form in just a moment.

A second 4-vector we can naturally make is for the source: we combine the vector current density with charge density ρ to make

$$J^\mu = (\rho c, J^x, J^y, J^z) .$$

Using this, we can write Maxwell's equations in a very simple form:

$$\partial_\nu F^{\mu\nu} = \frac{4\pi}{c} J^\mu , \qquad \partial^*_\mu F^{\mu\nu} = 0 .$$

Expand one example: The $F^{\mu\nu}$ equation for $\mu = 0$. Then,

$$\begin{aligned} & \frac{\partial F^{0\nu}}{\partial x^\nu} = \frac{4\pi}{c} J^0 \\ & \frac{\partial F^{0x}}{\partial x} + \frac{\partial F^{0y}}{\partial y} + \frac{\partial F^{0z}}{\partial z} = \frac{4\pi}{c}(c\rho) \\ & \frac{\partial E^x}{\partial x} + \frac{\partial E^y}{\partial y} + \frac{\partial E^z}{\partial z} = 4\pi\rho \\ \longrightarrow \qquad & \vec{\nabla} \cdot \vec{E} = 4\pi\rho \end{aligned}$$

If you expand and combine the $F^{\mu\nu}$ equation for $\mu = 1, 2, 3$, you get

$$\begin{aligned} -\frac{1}{c}\frac{\partial \vec{E}}{\partial t} + \vec{\nabla} \times \vec{B} &= \frac{4\pi}{c}\vec{J} \\ \longrightarrow \qquad \vec{\nabla} \times \vec{B} &= \frac{1}{c}\frac{\partial \vec{E}}{\partial t} + \frac{4\pi}{c}\vec{J} \,. \end{aligned}$$

Likewise, the $\mu = 0$ part of the ${}^*F^{\mu\nu}$ equation gives us

$$\vec{\nabla} \cdot \vec{B} = 0 \,;$$

the $\mu = 1, 2, 3$ parts give us

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \,.$$

24.4 General relativity

We've now got enough background that we can now sketch the general theory of relativity. Without going into too much detail the key quantity we need to describe gravity in a relativistic manner is the metric: our starting point is the equation

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \,.$$

Rather than being constant, as it was in special relativity, the metric is a varying function in general relativity. This turns out to mean that spacetime has *curvature*. All of gravity is encoded in the mathematics of this curvature.

The metric plays the same role for gravity that the 4-vector A^μ plays in electricity and magnetism — think of it as a "potential" for gravity. The equivalent of Maxwell's equations are the *Einstein Field Equations*:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \,.$$

Some explanation of this equation is in order. On the right-hand side, the tensor $T_{\mu\nu}$ — called the "stress-energy tensor" — describes the flow of 4-momentum. This tensor plays a role in these equations analogous to that of J^μ in Maxwell's equations. A component $T_{\mu\nu}$ describes the flow of the μ component of 4-momentum in the ν direction, in the same way that J^μ describes the flow of charge in the μ direction. For example, T_{00} describes the flow of the zero component of 4-momentum into the time direction; this turns out to be the energy density. Components like T_{0x} describe the flow of energy in the x direction. This is related to the mass density ρ_m and the velocity with which mass flows.

The tensor $G_{\mu\nu}$ is, in general, an enormous mess: it is found by applying a 2nd order, coupled nonlinear differential operator to the metric $g_{\mu\nu}$. The key thing to note is that, roughly speaking, it corresponds to *two* derivatives of the metric. Thus Einstein's equations are of the form "two derivatives of the potential (metric) equals the source", exactly as Maxwell's equations are.

24.4.1 Gravitational "forces"

Before moving on, we should talk about gravitational forces. In electricity and magnetism, the force that operates upon charges is

$$\vec{F}_{\text{E\&M}} = q\left[\vec{E} + (\vec{v}/c) \times \vec{B}\right] .$$

In general relativity, it turns out that there is *no such thing as a gravitational force!* Instead, what happens is that bodies move along a trajectory of *extremal time*. To understand what this means, go back to the definition of the invariant interval:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu .$$

For a body following a trajectory through spacetime, $ds^2 = -c^2 d\tau^2$, where τ describes the time as measured on a clock on that body. The path that the body follows through spacetime and the gravitational "forces" that act on it are found by working out the extremum of the total time τ that accumulates during its travels. The details of that calculation are somewhat beyond 8.022 — suffice it to say that it is a well-founded calculation.

An analogy helps to explain what is going on. Suppose we're on a flight from Boston to Tokyo. To save fuel, we want the trip to be as short as possible. The shortest path between two points is a straight line. But, on a round globe, what does "straight line" really mean? We go back to the old fashioned definition we learned in kindergarten: a "straight line" is the path which gives the shortest distance between two points. This is why a flight from Boston to Tokyo might actually go over the North Pole.

On a map, this flight might look crazy — it looks like you're bending the trajectory of the plane in as stupid a way possible. A person who believes in the reality of maps — as opposed to the reality of the earth — might say that the reason for this crazy trajectory is that "forces" act on the plane in such a way that the path over the North Pole is "easier" to follow than a simple logical "straight" line (straight according to the map, that is).

In general relativity, gravitational "forces" are like the fictional forces to which our map worshipper ascribes the airplane's trajectory.

24.5 A special metric

It turns out that the spacetime metric outside of a rotating body is given by

$$ds^2 = -(1 + 2\Phi)(c^2 dt^2) + (dx - \beta^x dt)^2 + (dy - \beta^y dt)^2 + (dz - \beta^z dt)^2 .$$

(Strictly speaking, this is an approximate result, only holding under certain conditions — namely that the body can't be too dense, and that its internal velocities must be much less than the speed of light.) The quantities β^x, β^y, and β^z are components of a vector $\vec{\beta}$ that has units of velocity.

If we run this metric through our procedure to find maximal time, we find that the "force" that acts on a body is given by

$$m\frac{d^2\vec{x}}{dt^2} = m\vec{g} + m(\vec{v}/c) \times \vec{H} ,$$

where

$$\vec{g} = -\vec{\nabla}\Phi , \qquad \vec{H} = \vec{\nabla} \times \vec{\beta} .$$

Notice that this force law looks *EXACTLY* like an electromagnetic force! In general relativity, the gravitational force looks just like the normal Newtonian force — encapsulated in the gravitational field $\vec{g}$ — is supplemented by a *gravitomagnetic* force arising from a field $\vec{H}$ that acts just like a magnetic field.

If we run this metric through Einstein's field equations, we find

$$\begin{aligned}
\vec{\nabla}\cdot\vec{g} &= -4\pi G\rho_m \qquad \text{(Gravitational Gauss's law)}\\
\vec{\nabla}\cdot\vec{H} &= 0\\
\vec{\nabla}\times\vec{g} &= 0\\
\vec{\nabla}\times\vec{H} &= -\frac{16\pi G\vec{J}_m}{c} \qquad \text{(Gravitational Ampere's law)}\\
\text{where} \qquad \vec{J}_m &= \rho_m\vec{v}\,.
\end{aligned}$$

The equations which fix the gravitational fields $\vec{g}$ and $\vec{H}$ look almost EXACTLY like Maxwell's equations! There are some important differences. Notice that the signs are reversed compared to the Maxwell equations. This is because gravity is *always* an attractive force. If the force between two masses were repulsive, we would find a + sign for the $\vec{\nabla}\cdot\vec{g}$. There is also a factor of 4 difference on the Ampere's law. This is harder to explain; it is related to the fact that the gravitational "potential" is a 2 index tensor, rather than a 1 index vector.

This analogy continues. If we move a gyroscope with an angular momentum vector $\vec{s}$ through this spacetime, it *precesses*: the vector feels a torque according to

$$\frac{d\vec{s}}{dt} = \frac{1}{2}\vec{s}\times\vec{H}$$

— an equation that is very similar to the torque felt by a magnetic dipole in a magnetic field, $\vec{\tau} = \vec{m}\times\vec{B}$.

These relativistic effects are ordinarily quite small and don't play much of an important role. Nonetheless, they are **not** zero. Last year, NASA launched a satellite to try to measure this "gravitomagnetic precession". This satellite, called *Gravity Probe B*[1] consists of an extremely precise and sensitive gyroscope that *very* slowly changes the direction of its angular momentum due to the gravitomagnetic field induced by earth's rotation. The net effect accumulated effect is about 41 milliarcseconds (1.1×10^{-5} degrees) *per year*. It's quite an amazing technological feat to be able to measure such a puny effect!

These effects play a much more robust role when gravitational fields and rotation speeds are stronger. In this case, the metric we used above is not really all that accurate. At least qualitatively, however, this picture of a Newtonian-like gravitational field $\vec{g}$ and a gravitomagnetic field $\vec{H}$, remains useful. The extra "force" arising from the gravitomagnetic fields are extremely strong and important! We look for such effects when we study objects in orbit around neutron stars and black holes — we *use* this gravitomagnetic effect as a means of learning about things in the universe.

[1]Gravity Probe A tested the fact that clocks run slow in gravitational fields.