PHYSICS 2415 E&M NOTES LECTURES 1 - 24

Michael Fowler, UVa

Introduction: What are these 2415 E&M Notes?

4/18/2025

I taught this course using PowerPoint, with demonstrations and clicker questions, etc. However, some students asked for traditional lecture notes to help review, etc. I've put together here the notes for the first 24 lectures, covering all e&m topics discussed except waves.

The first 9 lectures, almost a quarter of the course, are on electrostatics. Like the mechanical statics we covered last semester, this is about systems at rest, but now charges in equilibrium, with the total electrical force on each charge being zero. Actually, the book includes in this section motion of single electrons, or charged objects—what it doesn't include is the motion of huge numbers of electrons down a wire, an electric current—this is covered in the next four lectures.

The next 11 lectures cover magnetism, production of magnetic fields by electric currents, and the forces on electric currents from magnetic fields, the basis of electric motors and dynamos, going on to oscillating circuits, the basis of radio transmission.

1 Introducing Electrostatics: Electric Forces in Atoms, Molecules and Solids

The Atom and its Nucleus

Just to get going here: you all know, of course, that there are two kinds of electric charge, positive and negative, and that like charges repel, unlike charges attract. We'll discuss a little later how this is confirmed experimentally, and all the details. Now, this electric force is what holds atoms together by binding the electrons to the nucleus, and what holds atoms together to form molecules—so it's at the basis of all matter—if you take the atomic nucleus as given, which is fine for chemistry, biology, etc. The nucleus itself, though, is made of neutrons and protons, the protons are positively charged, so repel each other. The nucleus must therefore be held together by stronger force, which is termed, appropriately, the nuclear force. But if the nucleus is big enough, the repulsion between protons wins out, and it flies apart. This imposes a limit on the charge of a nucleus. The reason the electrostatic repulsion eventually dominates is that the nuclear binding force is strong but very short range, almost like a layer of glue on the particles, so only holds together protons and neutrons very close to each other, whereas the electrostatic repulsion is still there, even if lessened, at greater distances. So not only does it limit the size of nuclei, it's also the reason nuclear fission takes place.

In an atom, the nucleus contains Z protons each having charge e, Z is called the atomic number, and total positive charge Ze. The atom also has Z electrons, total charge $-Ze$. The electrons are much lighter than the protons and neutrons (about 1/2000) and they form a cloud surrounding the nucleus, of size around 0.1 nanometers, 10^{-10} meters. The nucleus is of order femtometers in size, 10^{-15} meters.

Just how the electrons are held in place cannot be explained without quantum mechanics. Even the size of the atom is inexplicable without the quantum. But in this course, we'll take it as given. Atoms can lose electrons if they are hit by other atoms or by strong electromagnetic fields. An atom missing one or more electrons is called an *ion*. It is positively charged, since the remaining electrons no longer balance the nuclear charge, so a sodium atom minus its electron is written Na^+. The electronic configurations are such that a sodium atom has one loosely attached electron, easy to lose. On the other hand, a chlorine atom has a gap in the electron pattern and easily accommodates an extra one. When lots of Na and Cl atoms are put together, it turns out that the Cl attracts the Na's loose electron so powerfully that it leaves home, settling into the Cl. The Na becomes Na^+, the Cl becoming Cl^-, then all these ions form a cubical pattern, with alternate atoms like a 3D chessboard, so each + has six – neighbors, etc., and electrostatic attraction keeps the whole thing together. (But it's not the whole story—there's repulsion between ions that get close. As we'll see later, electrostatic forces alone couldn't hold the ions stably in position. Quantum mechanics is essential.)

Yet if you put the salt into water, it dissolves! What happened to the electrostatic attraction? As we'll discuss later, water partially shields electric fields, it makes them weaker, and so ions near the surface of a crystal can be knocked off by thermal vibrations, and not return.

Conductors and Insulators

But not all solid formation can be explained in electrostatic terms: what if we just have a collection of sodium atoms? They will form a crystal too. It turns out that this time, the loose electrons leave their home atoms and wander throughout the crystal, so it can be visualized as a lattice of positively charged Na ions in a sea of negatively charged electrons. Why should this stay together? Unlike NaCl, where electrostatics does the job once you grant what happens to the atoms, here quantum mechanics is needed to make any progress in understanding—the uncertainly principle, and the electron spin, both quantum concepts, need to be put together with the electrostatic force to explain what's going on. We can't go into it at this point, but one result is important: this solid sodium has *electrons free to travel anywhere* in the crystal. A solid with such mobile electrons is called a *conductor* (of electricity) because the electric charge in it can move around. Good conductors are metals. They are shiny, because, as we'll see later, the mobile electrons reflect light well. They also conduct heat efficiently, again because it's the mobile electrons that carry the heat energy around.

In contrast, in solid salt the electrons are firmly pinned down to their atoms. Charge cannot move around, and—again for quantum mechanical reasons—any extra charge added finds it difficult to move around too. Solids (or fluids) with this property are called **i***nsulators*.

The transfer of electrons from Na atoms to Cl atoms, when they get close, has an analogy on a far larger scale: if you rub a glass rod with a piece of fur, both rod and fur become charged. (The technical name is the *triboelectric* effect—tribo just means friction.) Actually the *work* done in rubbing is irrelevant—the reason for rubbing is just to get the two surfaces in really close contact, so surface atoms on the two can transfer electrons. The glass gets positively charged. You can look up triboelectric series in Wikipedia to find out which way the charge moves—it's not obvious, rubbing glass with rabbit fur transfers electrons to the glass, cat fur has the opposite effect—at least according to some tables.

Charge Conservation is Always True

In an ordinary electrically neutral object, the total negative charge of the electrons exactly balances the total positive charge of the nuclei. Various experiments involving rubbing, or adding electric fields, etc., just move the electrons around, they do not change the number of electrons (they might become unbound, of course.) The total electric charge isn't changed. If we go for a moment beyond the energy range relevant for this course, in high energy (relativistic) collisions, electrons can disappear, but experimentally in such collisions either new negatively charged particles will be created, or positively charged particles will disappear: the *total* electric charge is *always* conserved.

Observing Electrostatic Forces

This is easy: if you run a rubber comb through your hair, it will become negatively charged, and will pick up small pieces of paper, which it wouldn't before. But wait a minute—the bits of paper weren't electrically charged! What happens is that when the charged comb is brought near, the electric field distorts the charge distribution within molecules in the paper. The electrons within molecules will move a bit away from the comb, leaving the positive charges slightly closer—even though no charge escapes from its own molecule. The net effect is that the positive charges, attracted by the comb, are slightly closer than the negative charges, which are repelled, so there is a net attraction.

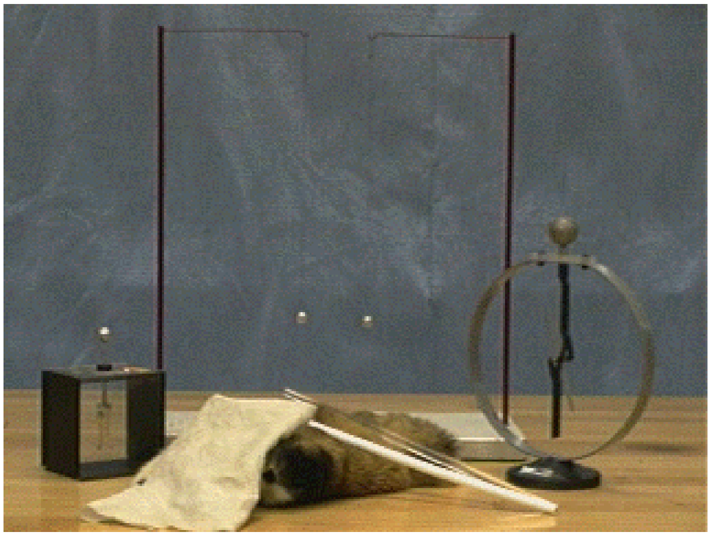
Little Charged Spheres

When we charge up the comb, or a glass rod, the charge is distributed in some complicated way on the surface of the rod. It can't spread itself evenly, since the rod is an insulator, so it's difficult to pin down exactly where the charge is. To make life simpler, we use small metal spheres, hung with insulating string (see the picture below). We rub the rod against the sphere, so some charge moves onto the sphere. Since the sphere is a conductor, the charge on it will spread evenly. If we charge two such spheres and hang them close to each other, but not touching, we see that they repel each other. Like charges repel—and the glass is positively charged. But if instead we use a Lucite rod, it's negatively charged on rubbing, and we can confirm that a positively charged sphere will attract a negatively charged one.

The Electroscope

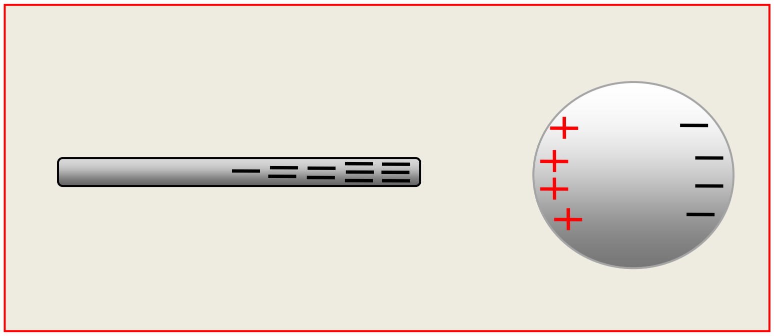
An electroscope detects the presence of charge by using repulsion of like charges. The simplest is the gold leaf electroscope: two very light leaves, of gold foil, hang down close to each other from a conductor. If charge is places on the conductor, it spreads onto the leaves, and they repel each other.

A slightly fancier version has a central column and a pivoting light rod, all made of metal, so if charge flows into the column and rod, the the repulsion causes the rod to swing away—this is the object on the right, the big circle, in the picture below.



Charging by Induction

If a negatively charged (by rubbing with silk) Lucite rod is brought close to, but not in contact with, a conductor, charge will redistribute on the conductor: the electrons will move away from the vicinity of the Lucite, that part of the conductor will therefore have net positive charge. But if we have two



conductors in contact, so charge can flow from one to the other, and the Lucite rod is brought close to one of them, that conductor will have a net positive charge. If, now, the two conductors are separated (by moving one, using some insulated contact) while the Lucite rod is close, the charge imbalance will of course remain even when the Lucite is removed, since the imbalanced charge will not be able to flow back. The Lucite rod never did touch the conductors: the electric charge on the rod itself did not move on to the conductors, but nevertheless the two conductors are now electrically charged, one positive the other equally negative. This process is called charging by induction.

The electroscope can be charged in this way: one puts one's finger on the sphere at the top, and brings, let's say, a positively charged glass rod close to the ball. This will attract electrons from the earth, through one's body and finger, on to the electroscope. If one now takes one's finger away, with the glass rod still in place, then removes the glass rod, the electroscope is now negatively charged, the needle will settle at some angle to the vertical. If now, the glass rod reapproaches, this angle decreases—because electrons are being drawn from the needle up to the ball.

2 Coulomb's Law, Field Lines

Coulomb's Law

Using the two small hanging spheres, we can even find just how the attraction varies with distance, by measuring the angle the string makes with the vertical and doing a simple calculation for varying distances. This is tricky, though—the charge slowly leaks away, especially in summer, moisture in the air dampens the surfaces slightly, and they conduct.

In an experiment essentially equivalent to this, Coulomb in the 1780's established that the electrostatic force decreased with distance as the inverse square, exactly like gravity (but of course it's much stronger!). He also found the force to be proportional to the magnitude of the charge. He accomplished that by using a charged sphere, then removing it and putting it in contact with an identical but uncharged sphere, so the charge would be equally shared. Now putting the sphere back, he found the force had been halved.

The Unit of Charge

In his honor, the unit of charge is called the coulomb. The charge on the electron is $1.6\text{x}10^{-19}$ coulombs. The coulomb is *a practical unit for dealing with batteries* and electric currents, it's the amount of charge flowing down a wire per second when the current is one ampere (we'll discuss these units in detail later). Unfortunately, though, in electrostatics we *never* deal with charge on this scale, and the microcoulomb is more typical.

For Electrostatics, this is an Immense Unit of Charge…

To picture the strength of electrostatic repulsion, imagine taking an ounce of water, and imagine you could pull all the electrons off the atoms, and put them is a separate glass. (Of course, this can't be done—as you'll see!) Now put our glass of electrons and your glass of nuclei one thousand kilometers apart, say here and Orlando. What's the strength of the attractive force between them?

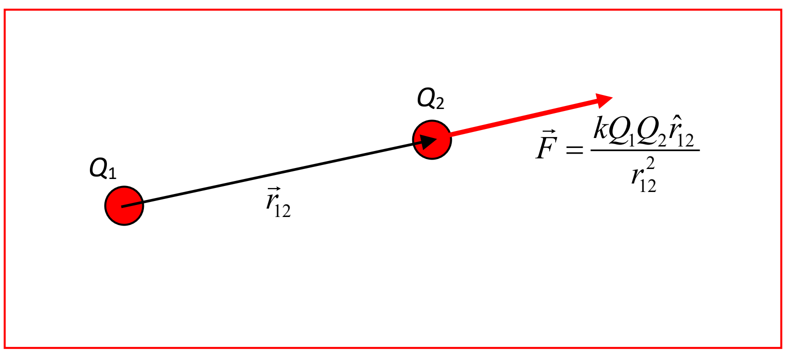
Now Avogadro's number of molecules, $6\text{X}10^{23}$, weigh a gram mole, that's 18 grams for water, so an ounce, 28 grams, is about 10^{24} atoms, or 10^{25} electrons (a water molecule has ten electrons total). One electron has charge $1.6\text{x}10^{-19}$ coulombs, so we have $Q = 1.6\text{x}10^{6}$ coulombs. Here to Orlando $r = 10^{6}$ meters, so

$$F = kQQ/r^2 = 9\text{x}10^9\text{x}(1.6\text{x}10^6)^2/10^{12}\text{, about } 2\text{x } 10^{10}\text{N … 2,000,000 tons.}$$

From this, we can definitely conclude that for charged spheres repelling each other, the *imbalance* in electron numbers from neutral is very, very small: this is why typical electrostatic charges are *micro*coulombs, but *total* electron charge in a sphere is of order *mega*coulombs—the imbalance is of order 10^{-12} or so.

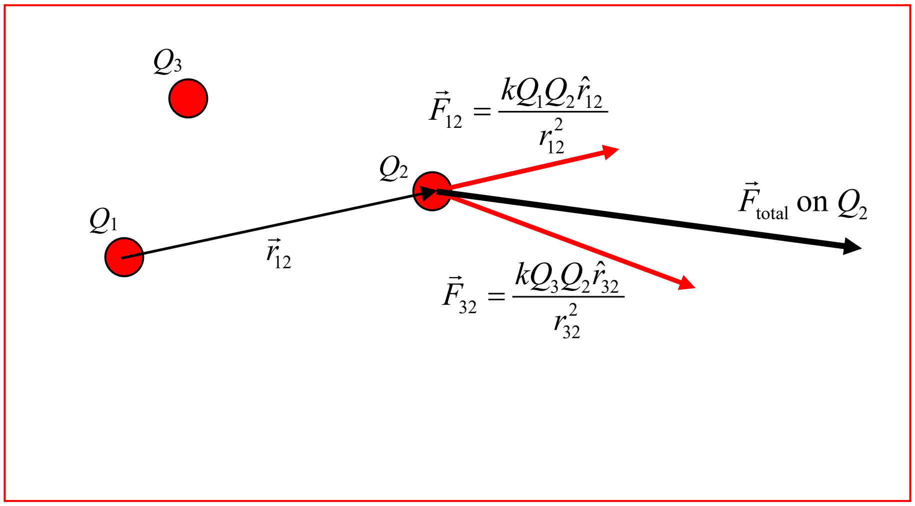
Coulomb's Law in Vector Form

We follow standard practice in denoting a vector of unit length parallel to the vector $\vec{r}$ by $\hat{r}$:



The Principle of Superposition

Electric force vectors *add*: if charge Q_2 is repelled by charge Q_1 and charge Q_3, the total repulsive force on it is the vector sum of the separate repulsive forces:



This may look obvious, but must be experimentally verified—the Law of Superposition is in fact not true for nuclear forces!

The Electric Field

Just why a charge can affect the motion of another charge some distance away is rather mysterious, as indeed is the gravitational attraction between two masses. Einstein was the first to realize that a mass distorts space time in such a way that other masses, instead of moving at constant velocity, accelerate. The earth's gravitational field slightly distorts space and time here so all masses free to fall accelerate downwards at the same rate (ignoring other forces, such as air resistance, of course).

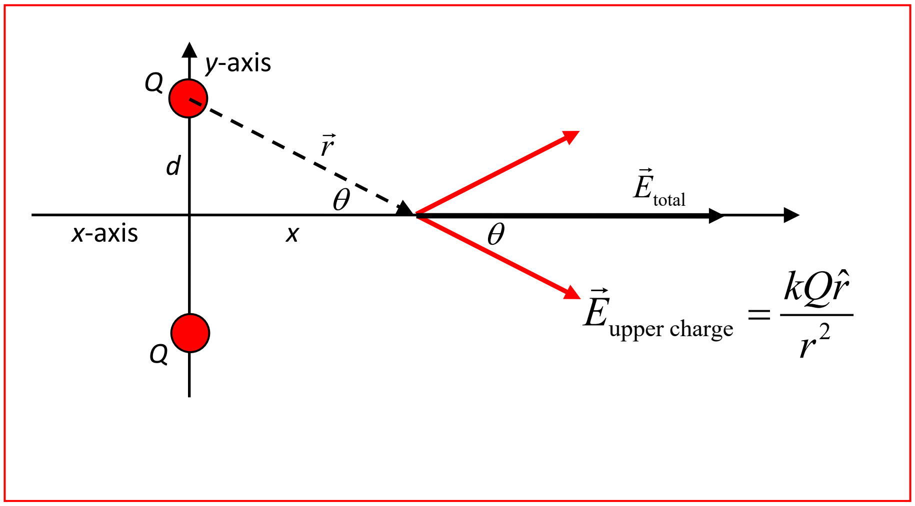
An electric charge does not distort spacetime, but does have a surrounding energy density in space, called its electric field. Another charge placed in this field experiences the inverse-square force.

The electric field $\vec{E}(\vec{r})$ *at a point* $\vec{r}$ *is defined by stipulating that the force* $\vec{F}$ *on a tiny test charge q at* $\vec{r}$ *is equal to* $q\vec{E}$ *.*

Therefore, the electric field at a point can be determined experimentally without knowing the details of the placement of charges producing the field. It may not be necessary to know that. The reason for taking a small test charge is that if the field is partially from charges on conductors, introducing a large test charge will change the distribution of the charges on conductors, so changing the field being measured. However, if the field is from point charges, or charges on insulators, any size test charge will be fine.

Field from Two Equal Charges

Two charges *Q* are placed on the *y*-axis, equal distances *d* from the origin up and down. What is the electric field on the *x*-axis, and where does it reach a maximum value?



It's clear from the diagram that at any point on the *x*-axis, the sum of the electric fields from the two charges is itself along the axis, and has value

$$E = \frac{2kQ}{r^2}\cos\theta = \frac{2kQx}{r^3} = \frac{2kQx}{\left(x^2+d^2\right)^{3/2}}.$$

Notice that at large distances (*x* >> *d*) this goes to $2kQ/x^2$, the same as a charge 2*Q* at the origin; but actually at the origin the electric field is zero: the two components are equal and opposite. For small *x*, the field strength increases linearly with *x*. Clearly, if the field with increasing x first increases then finally decreases, it must have a maximum value somewhere. To find where this is, we use *dE*/*dx* = 0 at that point. Routine differentiation of the above expression gives the value $x_{E_{\max}} = d/\sqrt{2}$.

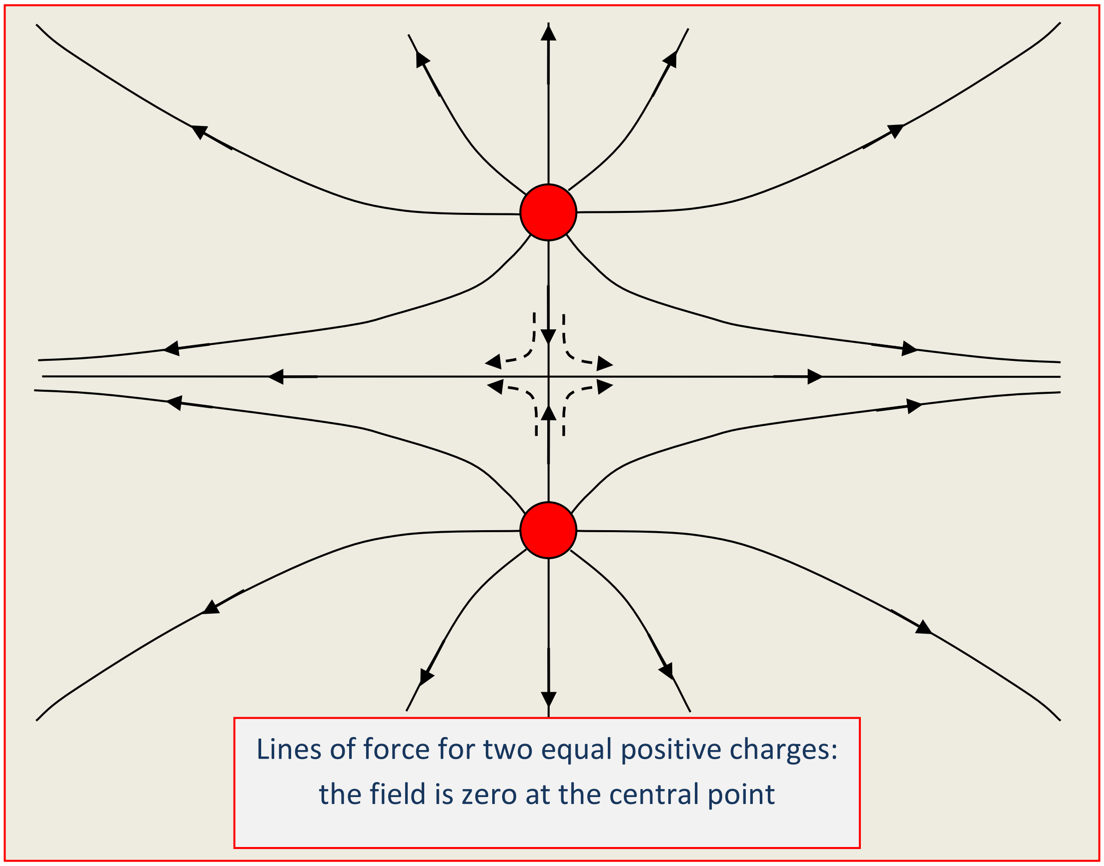
Lines of Force

One way to visualize the electric field is to draw lines of force. These are lines drawn so that at any point on the line, the electric force on a positive test charge is in the direction of the line.

Exercise: Try sketching the lines of force for the two equal charges in the diagram above.

We already know the *x*-axis is a line of force, since the field everywhere on it is along the axis, but in different directions for positive and negative *x*. We also know that anywhere on the *y*-axis, the force is along the *y*-axis, pointing away from the nearer charge. We know that close to one of the charges, its force will dominate, so the field lines initially come out close to radially from the charge. Finally, far away the two charges look like one, so again the field lines will be radial at large distances.

Here's a sketch:



Lines of force for two equal positive charges: the field is zero at the central point

To get a better idea of electric fields in this and other cases, go to [this website](). Sometimes attempts are made to indicate the strength of the field by how close together the lines are drawn. This is easy to see for the simplest case of a single charge, although even there one would think from a diagram on paper that the field was decreasing as $1/r$, because that's how the lines thin out—but really the field is in *three* dimensions, so actually they would thin out faster, as $1/r^2$, in a (more realistic) 3D model. The other problem with this approach is that where the field is weak, no lines appear at all, so it's difficult to figure out what's going on. We've added some dotted local lines of force near the midpoint, the place where the field strength goes to zero.

Field on the Axis of a Ring of Charge

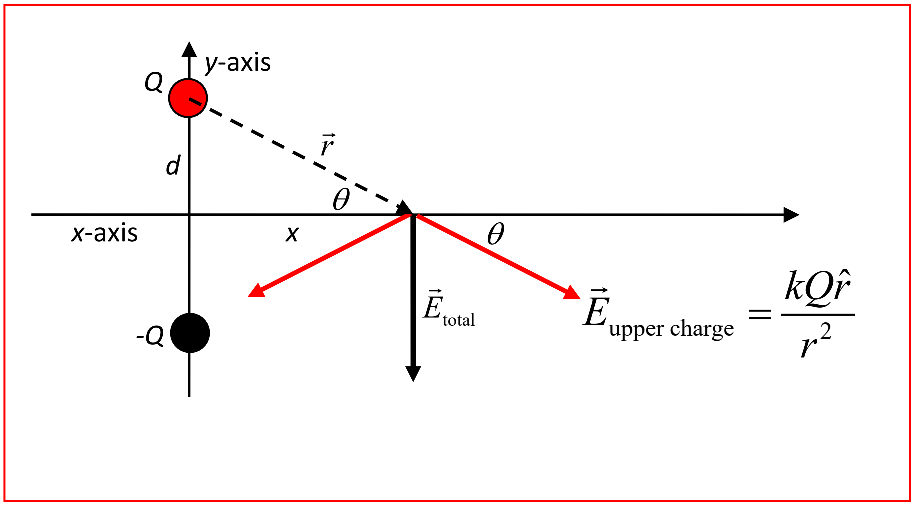
Given what we've just done, this is very easy: if a ring has total charge Q, uniformly spread around the ring, it can be replaced by a large number of small pairs of charges arranged as in the above example, and the axis of the ring is the x-axis in the above picture for all these pairs of charges, the electric fields of all the pairs add, they're all along the axis for a point on the axis, so the answer has to be, for a ring of radius d,

$$E = \frac{kQx}{\left(x^2 + d^2\right)^{3/2}}.$$

3 Dipole, Charged Line

The Dipole

Suppose now that in the previous example we replace the lower charge by –*Q*:



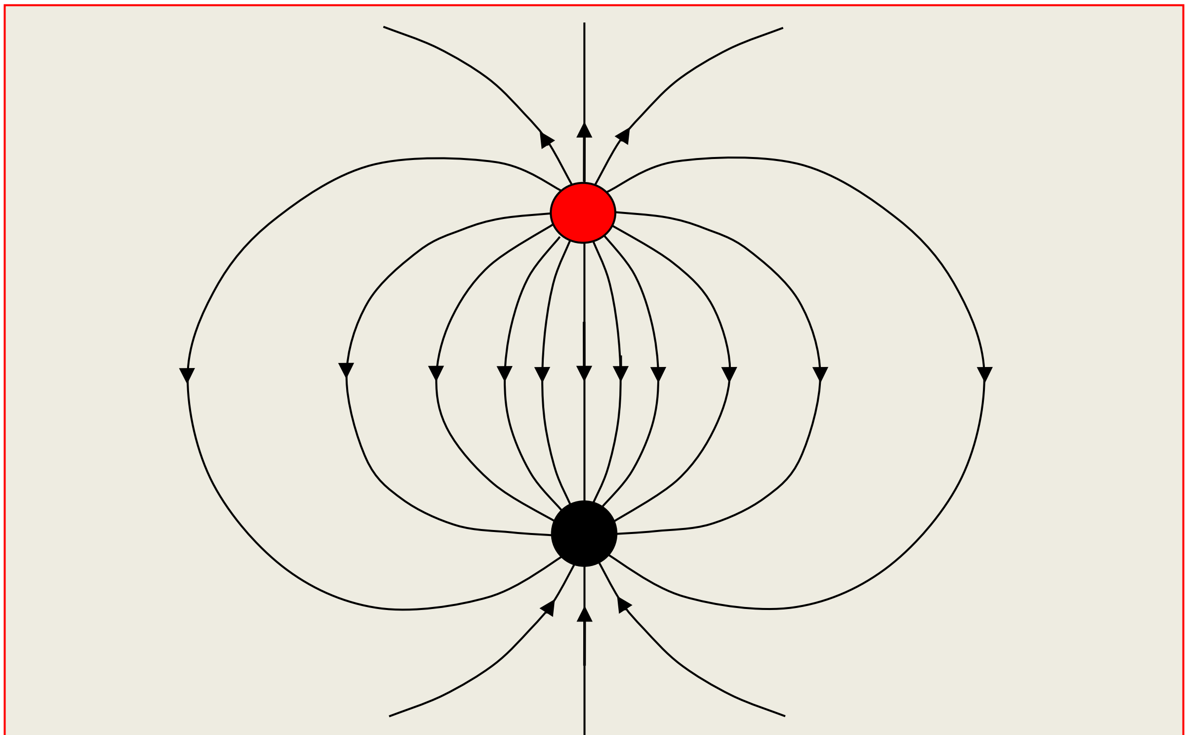
Notice that now the electric field is perpendicular to the *x*=axis. It has magnitude

$$E = \frac{2kQ}{r^2}\sin\theta = \frac{2kQd}{r^3} = \frac{2kQd}{\left(x^2+d^2\right)^{3/2}}.$$

On the *x*-axis, this force is maximum at the origin, the point midway between the two charges, and at large distances it decreases as r^{-3}, because the inverse square repulsion and attraction are opposing each other.

A pair of equal and opposite charges close together like this is called a **dipole**.

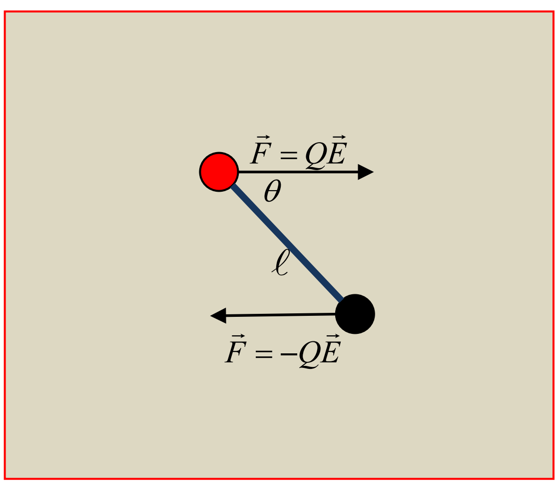
The electric field lines radiate outwards from the positive charge, inwards to the negative charge, and must cross the *x*-axis at right angles, giving the following picture:



Electric field lines of force for a **dipole**: charges *Q*, -*Q* a fixed distance ℓ apart

Force on a Dipole in an External Electric Field

Suppose an electric dipole (imagine now that the two charges are connected by a light insulating rod) is placed in an electric field of uniform strength, so its positive and negative charges feel forces $Q\vec{E}$ which are equal in strength, but of course opposite in direction. There is then no net force on the dipole, but there is in general a torque,

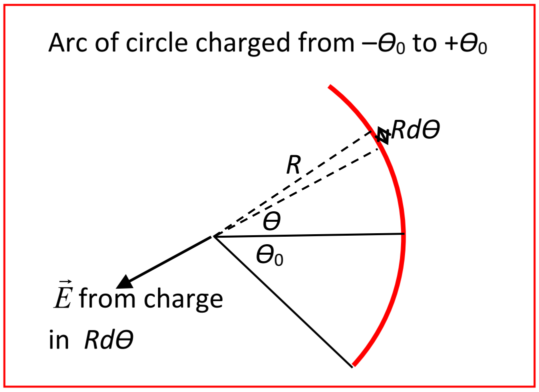


$$\vec{\tau} = \vec{p} \times \vec{E},$$

where the vector dipole moment $\vec{p}$ has magnitude $Q\ell$, and direction along the line of the dipole from the negative charge to the positive charge.

Electric Field from a Continuous Charge Distribution

So far, we've looked at fields from a few charges, apart from the field along the axis from a uniform ring of charge. What if the charge is only on part of the ring? Let's find the field at the center from a uniformly charged arc of a circle, radius *R*, with charge density λ coulombs/meter from $-\theta_0$ to $+\theta_0$.



The strategy is to find the electric field from one small length of the arc, then add all the small arcs—in other words, do an integral.

The length of arc between θ and $\theta + d\theta$ has length $Rd\theta$ and hence charge $\lambda Rd\theta$, so it gives a contribution to the electric field at the center of the circle of magnitude $k\lambda Rd\theta / R^2 = k\lambda d\theta / R$ at an angle θ to the *x*-axis. From the symmetry of the problem, the total field must be along the axis, so we only need count the component $(k\lambda / R)\cos\theta d\theta$ in that direction: evidently the total field from the whole arc has magnitude

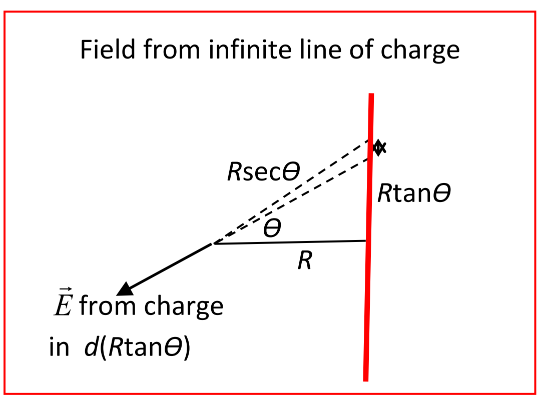
$$E = (k\lambda / R) 2\sin\theta_0$$

and points in the negative *x*-direction.

It's worth noticing that this field only goes down as 1/*R*, not 1/*R*^{2}: if the radius is doubled, the distance is doubled but so is the amount of charge. We'll see this 1/*R* behavior in the next paragraph for an infinite *line* of charge, for a similar reason.

Field from a Uniformly Charged Infinite Line

This is a slightly more complicated version of the problem above: now *r* varies, and θ goes from $-\pi/2$ to $\pi/2$. Again, from the symmetry, the total field must be directly outwards from the infinite line of charge, so we need only calculate the component in that direction.



For uniform line density of charge λ coulombs/meter, the amount of charge corresponding to a small angle $d\theta$ as shown in the diagram is $\lambda dy = \lambda d(R\tan\theta) = \lambda R\sec^2\theta d\theta$. This is at a distance $R\sec\theta$ from the origin (the point where we're finding the field), so contributes an electric field there of strength $kq/r^2 = k\lambda R\sec^2\theta d\theta / R^2\sec^2\theta = k\lambda d\theta / R$.

This must be multiplied by $\cos\theta$ to give the *x*-direction component, then integrated from $-\pi/2$ to $\pi/2$ to give the total field:

$$E = \frac{k\lambda}{R}\int_{-\pi/2}^{\pi/2}\cos\theta d\theta = \frac{2k\lambda}{R}$$

And of course it points directly away from the (positively charged) wire.

If the wire is not infinite, the limits on the above integral are changed appropriately, and there is also an electric field component parallel to the wire, except at a point level with the center. This component can be found by integrating in the same way. Try drawing the lines of force for this case.

Field from a Uniformly Charged Infinite Plane

Think of the infinite plane of charge as made up of a huge number of lines of charge parallel to the *y*-axis, each having charge density λ coulombs per meter, and along the *x*-axis there is a line density of μ of these lines per meter. This means that there is a area density of charge in the plane of $\sigma = \lambda\mu$ coulombs per square meter.

Now, imagine that in the diagram above we used to calculate the field from an infinite line of charge, each bit of the charge line now represents lines of charge perpendicular to the paper—this is how we replace the line of charge by a sheet of charge. Recall the electric field strength from charge *q* was kq/r^2, that from a line is $2k\lambda/r$. To find the field from the lines in the angle $d\theta$, we now use the line density μ in place of λ, so the number of lines in $d\theta$ is $\mu dy = \mu d\left(R\tan\theta\right) = \mu R\sec^2\theta d\theta$ and their contribution to the electric field is $\dfrac{2k\lambda}{R\sec\theta}\mu R\sec^2\theta d\theta$.

The total electric field from the uniformly charged plane is therefore, taking the *x*-components:

$$\int_{-\pi/2}^{\pi/2}\frac{2k\lambda}{R\sec\theta}\mu R\sec^2\theta\cos\theta d\theta = 2\pi k\lambda\mu = \frac{\sigma}{2\varepsilon_0}$$

recalling that $\sigma = \lambda\mu$ and $k = 1/4\pi\varepsilon_0$.

So the electric field goes out from an infinite uniform sheet of charge perpendicular to it, and does not decrease with distance. For a finite sheet, this picture is still good for distances small compared with the size of the sheet.

A common situation (for example, inside a capacitor) is to have two parallel uniform sheets of charge, of equal charge density, but one positive and the other negative. The distance between the sheets is much less than their size, so it's a good approximation to take them as infinite. The electric field is then easily found by superposition: it is a uniform σ/ε_0 between the sheets, and zero outside.

Electric Fields and Conductors

By definition, in a conductor there are charges that are free to move. If there is an electric field, they will move. In electrostatics, all charges are at rest. Therefore, in an electrostatic situation the electric field inside every conductor must be zero. It also follows that the lines of force of the electric field outside the conductor must approach the surface perpendicularly, because if there were a parallel component, a current would flow in the surface.

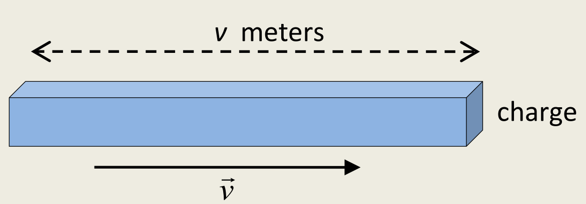
home

4 Gauss' Law

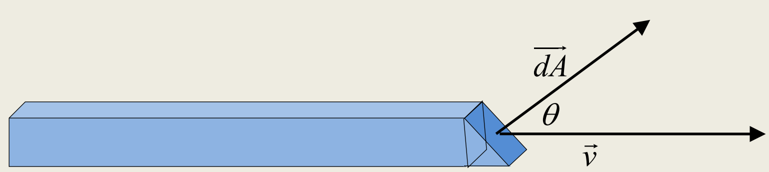
Electric Flux: a Watery Analogy

The main concept in Gauss' Law is *electric flux*. What does this mean? The word *flux* just means *flow,* for example an influx of people into a room means they're coming in. Before talking about electric flux, let's look at something easier to visualize: flow of water.

We'll begin by considering flow down a river. Suppose we stretch a net across the river, a fisherman's net with thin strings and approximately square small holes, so that all the water flowing down the river goes through the net. For a steadily flowing river, the total flow through the net, in, say, cubic meters of water per second, doesn't depend on whether the net is stretched flat across the river, or is curved by the current so that it bulges in a downstream direction—in either case, the total flow is all the water in the river. (I'm assuming here that the strings themselves are thin enough not to affect the flow measurably.)



If a fluid is flowing steadily at *v* meters/sec, the volume through an area *dA* perpendicular to the flow is *vdA* m^3/sec.



If the fluid is flowing through a small area *dA* having its normal at an angle θ to the direction of flow, the effective area seen by the flow is $dA\cos\theta$, the volume through the area is $vdA\cos\theta = \vec{v}\cdot\vec{dA}$ m^3/sec.

One way to find the total flow is to add the flows through all the little squares. We'll assume the squares are small enough that the fluid velocity doesn't vary significantly over one square: first, assume the little square is perpendicular to the direction of flow: if the square has area *dA* square meters (it's small), and the fluid is flowing at speed *v* meters per second, then in one second a volume *vdA* cubic meters of fluid will flow through the square. But what if the square is *not* perpendicular to the flow? Then what counts is the *effective area* the flow sees—if the normal to the square is at an angle θ to the flow, this effective area is $dA\cos\theta$.

The standard notation is to represent the area by a vector $\overrightarrow{dA}$ of magnitude dA, and direction perpendicular to the area, that is, along the normal. (The sign is of course ambiguous—we have to decide which way it's pointing on a case by case basis.)

Then the flow across the small area $\overrightarrow{dA}$ is $\vec{v}\cdot\overrightarrow{dA}$ (we have now chosen the vector $\overrightarrow{dA}$ to point downstream), and the total river flow F through all the holes in the net is

$$F=\int\vec{v}\cdot\overrightarrow{dA}.$$

It is important to realize that this total flow cannot depend on the detailed shape of the net: it must be the same for all nets that completely span the river, so that all the river water passes through the net.

Flow from a Point Source

To take a slightly different example, consider filling a large deep swimming pool using a hose, the end of the hose being deep in the water. We'll assume there are no currents present except the water flowing out of the end of the hose, and that this outflowing water goes out equally in all directions: this could be achieved, for example, by covering the end of the hose with a porous ball, so the water flows directly outwards from this ball (we'll ignore the obstruction presented by the incoming hose itself—suppose it's really thin).

Imagine now surrounding the source with a fishnet bag, a complete surface surrounding it, so all the water coming out the source goes through some hole in this fishnet. It's easy to see that if the hose is delivering *F* cubic meters per second, this will also be the total flow through the fishnet in a steady situation—water is not going to pile up inside the bag, it's incompressible for all practical purposes.

That is,

$$F=\int\vec{v}\cdot\overrightarrow{dA}$$

and this integral is the same for any closed surface surrounding the source.

At this point, we'll abandon the fishnet picture, and talk a little more abstractly about integrating over a surface surrounding the source, with the increment of area denoted by $\overrightarrow{dA}$ pointing outwards.

In particular, let's take a spherical surface of radius r surrounding the source. Since we've said the water is flowing out symmetrically in all directions, it will have the same speed $v(r)$ at all points on this centered spherical surface, and the flow vector will be parallel to the normal to the surface, so the total flow

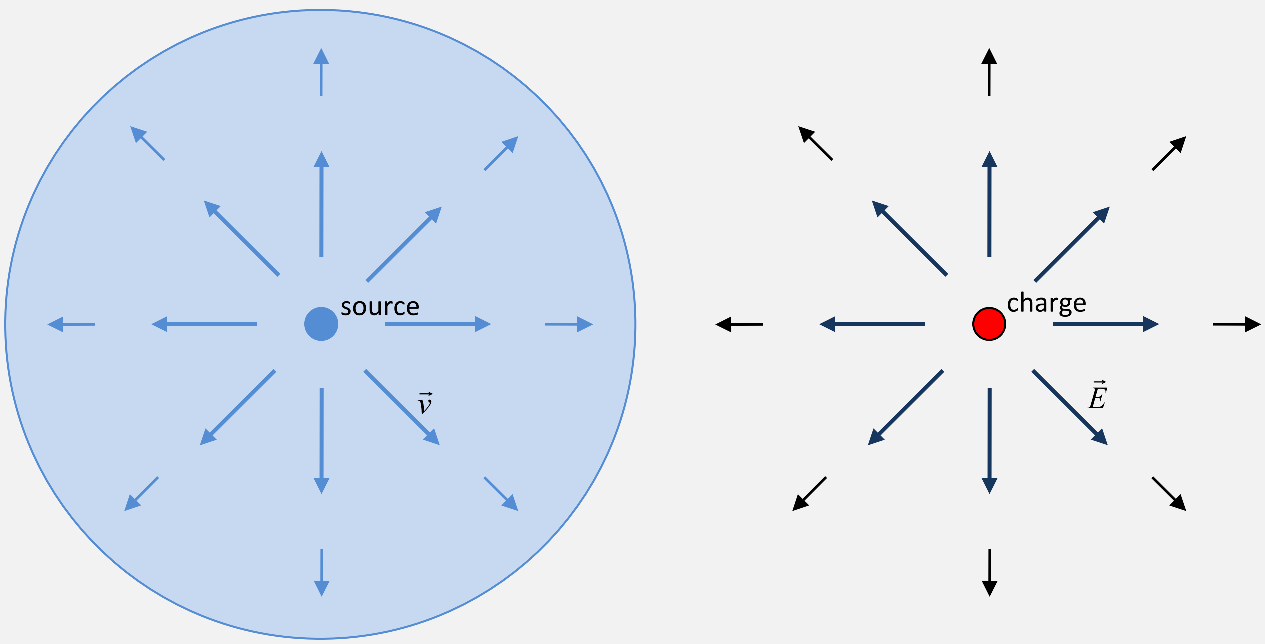
$$F = \int \vec{v} \cdot \overrightarrow{dA} = \int v dA = v \int dA = 4\pi r^2 v.$$

Therefore,

$$\vec{v}\left(\vec{r}\right) = \frac{F}{4\pi r^2}\hat{r}.$$

Notice this is formally identical to the electric field from a point source:

$$\vec{E}\left(\vec{r}\right) = \frac{Q}{4\pi \varepsilon_0 r^2}\hat{r}.$$



For water flowing equally outwards in all directions from an underwater source, the vectors $\vec{v}$ denoting water velocity at each point in space form a pattern *identical* to that of the electric field vectors $\vec{E}$ from a point charge.

Note: Remember also this picture is a ***2D*** representation of a ***3D*** reality!

This is why historically people talked about "electric flux": the electric field vector from a point charge looks exactly like the fluid velocity vector for an incompressible fluid flowing symmetrically outwards from a small spherical source.

Specifically, the electric flux through a small area is defined in exact analogy with the flow of fluid through an area, it is just $\vec{E}\cdot\overrightarrow{dA}$, and the total electric flux through a closed surface with a single charge inside it is given by

$$\int \vec{E}\cdot\overrightarrow{dA} = Q/\varepsilon_0.$$

We know the integral doesn't depend on which enclosing surface we choose, because the electric field vector is everywhere proportional to our water flow vector. We know the constant is Q/ε_0 because that's what we get if we take a spherical surface, with the charge at the center:

$$\int \vec{E}\cdot\overrightarrow{dA} = \int \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{r}\cdot\overrightarrow{dA} = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\int dA = \frac{Q}{\varepsilon_0}.$$

(The outward pointing unit vector $\hat{r}$ is parallel to the outward pointing little area vector $\overrightarrow{dA}$.)

We should mention that for a negative charge, the field lines of course point inwards: the fluid analogy is sucking the water out of the pool, a drain point.

What if we have a closed surface that doesn't include our point charge? What is $\int \vec{E}\cdot\overrightarrow{dA}$ in that case? The answer should be obvious from the flowing water analogy: if there is no source of water inside the surface, the water flowing in must balance the water flowing out in the steady state. That is to say, if there is no charge inside a surface in an electrostatic problem, then $\int \vec{E}\cdot\overrightarrow{dA} = 0$.

Gauss' Law for General Charge Distributions: Use Superposition!

We've given a detailed account of the value of $\int \vec{E}\cdot\overrightarrow{dA}$ over a closed surface for the field from a single charge, it's equal to Q/ε_0 if the charge is inside the surface, zero otherwise.

But it's easy to generalize, because any charge distribution can be represented as a (possibly large) number of point charges, and *the total electric field is the linear sum* of all the electric fields from these many point charges:

$$\vec{E}(\vec{r}) = \vec{E}_1(\vec{r}) + \vec{E}_2(\vec{r}) + \vec{E}_3(\vec{r}) + \vec{E}_4(\vec{r}) + \dots$$

and therefore for a closed surface

$$\int \vec{E}\cdot\overrightarrow{dA} = \int \vec{E}_1\cdot\overrightarrow{dA} + \int \vec{E}_2\cdot\overrightarrow{dA} + \int \vec{E}_3\cdot\overrightarrow{dA} + \int \vec{E}_4\cdot\overrightarrow{dA} + \dots$$

The first integral in the series will equal Q_1/ε_0 if the charge Q_1 is inside the closed surface, zero otherwise. The same is true for all the terms in the series, so we conclude:

$$\int_{\text{closed surface}} \vec{E}\cdot\overrightarrow{dA} = \left(\text{total charge inside surface}\right)/\varepsilon_0$$

and this is *Gauss' Theorem.*

5 Using Gauss' Theorem: Spheres, Lines, Planes

Taking Advantage of Symmetry

In general, integrating a vector flux over a surface is a daunting task, but in certain symmetric cases it's very easy, and can then be used to find electric fields much more easily than by adding contributions from large (or infinite, in the case of a continuous distribution) numbers of separate charges.

Spherical Shell

A good example is finding the electric field from a uniformly charged spherical shell, say charge *Q* and radius *R*. Since the sphere is uniformly charged, it has perfect spherical symmetry, it is not altered by turning the sphere through some angle. Therefore, the electric field must also be spherically symmetric. The *only* spherically symmetric electric field has the field pointing directly outwards (or inwards) from the center at all points.

Let's apply $\int_{\text{closed surface}} \vec{E}\cdot\overrightarrow{dA} = \left(\text{total charge inside surface}\right)/\varepsilon_0$

to a spherical surface of radius *r* bigger than the sphere of charge, but with the same center.

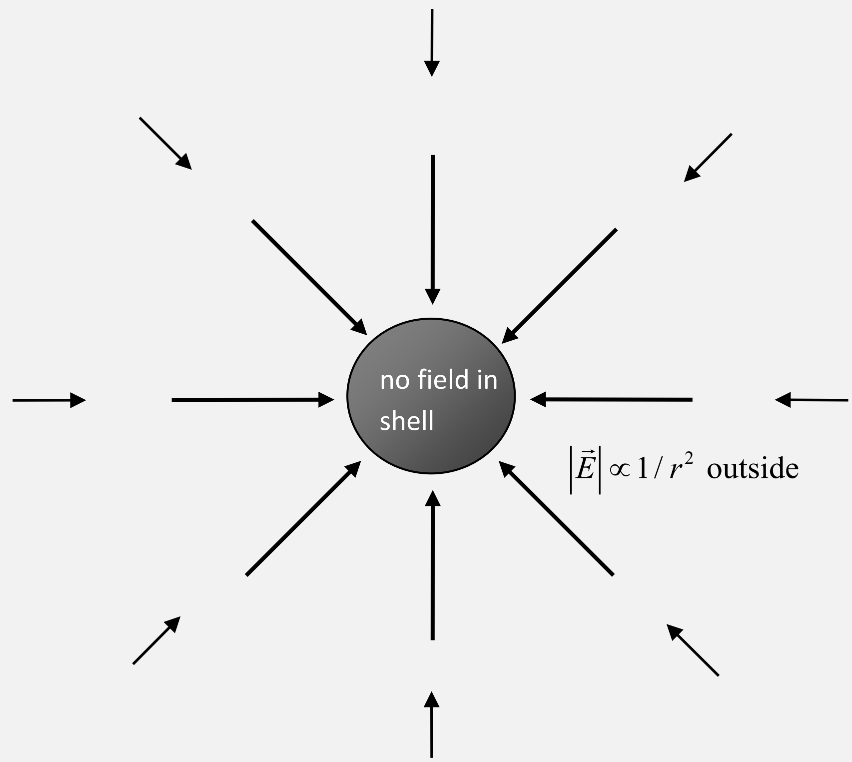
The field $\vec{E}$ points outwards everywhere on the surface, so it's parallel to $\overrightarrow{dA}$, and has the same strength everywhere on the sphere, by symmetry. The total area of the sphere is $4\pi r^2$, so the integral is equal to $4\pi r^2 E$, and outside the sphere of charge:

$$\vec{E}(\vec{r}) = \frac{Q}{4\pi\varepsilon_0 r^2}\hat{r}$$

the same as for a point charge at the center. It's worth mentioning that since gravity is also an inverse square force, this same result is true for the gravitational field from a spherical shell of mass. (This can be proved using Coulomb's Law or its gravitational equivalent, but it's quite difficult—it's done here.)

What about the electric field inside the sphere? We do the same trick: integrate over a spherical surface with the same center as the sphere of charge. This time, though, there is no charge inside our smaller spherical surface, so *the electric field must be exactly zero inside the sphere*.

The complete picture of the electric field for a uniformly charged shell is therefore:



Electric field vectors for a uniformly **negatively** charged spherical shell

Solid Sphere

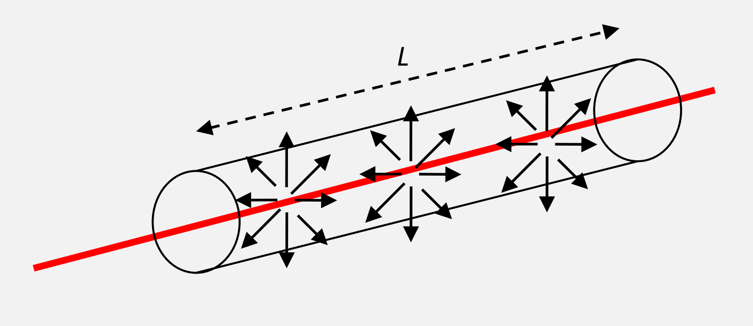
The key for any spherically symmetric charge distribution is superposition: the distribution can be expressed as the sum of (or integral over) spherical shells. The contribution from each shell is zero inside that shell, and equal to that from a point charge at the center outside the shell. So, for the case of a uniformly charged (throughout the volume) sphere, outside the whole sphere the field is the same as if all the charge were at the center, inside the solid sphere, at distance *r* from the center, it's the same field as from a point charge at the center equal to the amount of charge in a sphere of radius r: in other words, there is no contribution from those shells the point is inside.

This uniformly charged sphere is not a likely object to find in electrostatics, but it is exactly equivalent to the gravitational field for a sphere of uniform density, a much more realistic problem. And, in fact, the electrostatic uniformly charged sphere was a subject of intense interest a century ago, as a possible model for the atom: before the nucleus was discovered, but it was already known that the atom contained negatively charged electrons, it was suggested that the positive charge was spread over a sphere, and the electrons were inside this sphere: this was called the "plum pudding" model.

For a nonuniform spherical distribution, the same approach works: the field at any point is equivalent to a point charge at the center equal to all the charge between the point and the center.

Lines and Cylinders of Charge

Gauss' theorem works well for finding the electric field from an infinite uniform line of charge. From symmetry, the field lines must be directed perpendicularly to the line of charge, and the field strength can only depend on distance from the wire. For our Gaussian surface, we take a cylinder of length one meter and radius *r*, the wire running along the axis of the cylinder.



Electric field vectors and Gaussian surface of integration for a positively charged infinite straight wire

The total area of the cylinder is $2\pi r$ so, using

$$\int_{\text{surface}} \vec{E}.\overrightarrow{dA} = (\text{enclosed charge})/\varepsilon_0$$

the enclosed charge,

$$2\pi rLE(r) = \lambda L / \varepsilon_0$$

from which

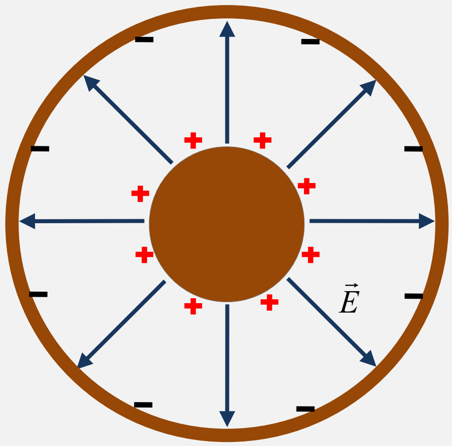
$$E(r) = \frac{\lambda}{2\pi r\varepsilon_0}.$$

(Easier than using Coulomb's Law for the field from each increment of charge and integrating!)

This same method applies for finding the electric field from a uniformly charged cylinder of charge. Just imagine the wire in the picture above being replaced by a fatter wire, then by a hollow cylinder, but staying inside the Gaussian cylindrical surface we integrate over. We get the identical result for $E(r)$, now we must interpret λ as the charge on one meter of the whole cylinder. If this is a hollow cylinder, a pipe, taking a Gaussian surface inside it, the surface encloses no charge, so the electric field inside a hollow cylinder from the charge on the cylinder is zero.

Coaxial Cable

Of course, we could add a line of charge, or even another cylinder, inside our charged cylinder, in which case the total electric field would be the sum of the electric fields from the two cylinders, using superposition. In fact, this is a coaxial cable, the cable used to transmit TV signals. etc. A coaxial cable (the word means "same axis") has a central copper wire, inside a hollow copper cylinder (see figure below). Between the two is a nonconducting dielectric—we'll discuss dielectrics shortly. The transmission of electromagnetic waves, the TV signal, is of course not an electrostatic situation, but nevertheless Gauss' Law still holds, and at any moment there are equal amounts of charge per unit length of cylinder on the surface of the central wire and the inner surface of the cylinder, and consequently an electric field as shown (there are also currents in the copper producing magnetic fields—more about that later).

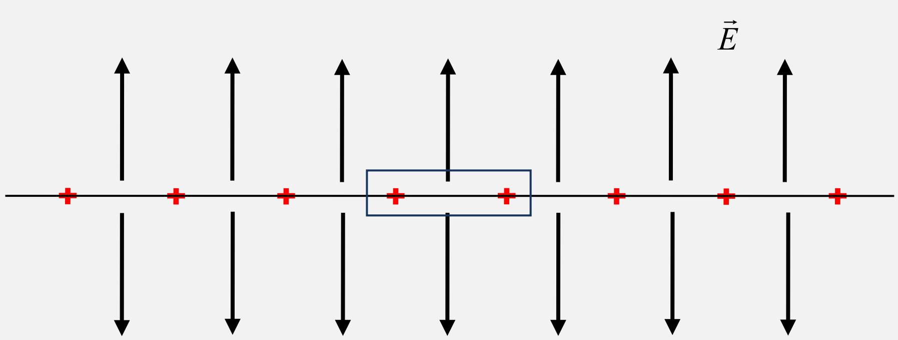


Typical electric field configuration in a coaxial cable, usually a copper cylinder and a central copper wire. The charge is on the *outside* surface of the *inner* conductor, the inside surface of the outer conductor.

The electromagnetic fields are entirely contained inside the cable, in contrast to signals sent down a pair of wires, and the outer cylinder protects the signal from external interference (and makes eavesdropping more difficult—you'll have to cut into the cable).

Uniformly Charged Plane

Gauss' Law makes it extremely easy to find the electric field from a uniformly charged plane, in contrast to the tedious integration necessary using Coulomb's law to find the electric field from each little area of the plane and taking the sum.



Electric field from a uniform plane of charge

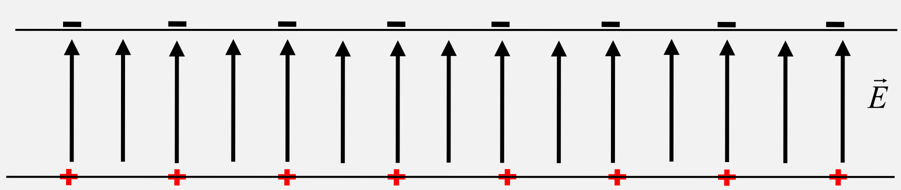
From symmetry, taking the plane to have infinite extent, the field must be perpendicular to the plane as shown above, where the plane of charge is seen in cross section, that is, the plane is perpendicular to the paper. Of course, the charge is distributed uniformly over the plane, with area density σ coul/m^2. To use Gauss' Law, we choose a surface shaped like a pillbox, represented in cross section by the rectangle above. The top and bottom surfaces both have area A, and an area A of the charged plane is included. The electric field is parallel to the area vectors on both the top and bottom surfaces, so the total contribution from those surfaces to $\int \vec{E} \cdot \overrightarrow{dA} = 2EA$. There is no contribution to the integral from the sides of the pillbox, as the electric field is parallel to those sides, thus $\vec{E} \cdot \overrightarrow{dA}$ is zero there. It follows immediately that

$$\int_{\text{surface}} \vec{E} \cdot \overrightarrow{dA} = 2EA = A\sigma / \varepsilon_0, \quad \text{so} \ \ E = \sigma / 2\varepsilon_0.$$

For an actual physical finite plane of charge, this value of E is a good approximation at points close to the surface relative to the size of the plane. For distances large compared to the extent of the plane, the field becomes more like that from a point charge.

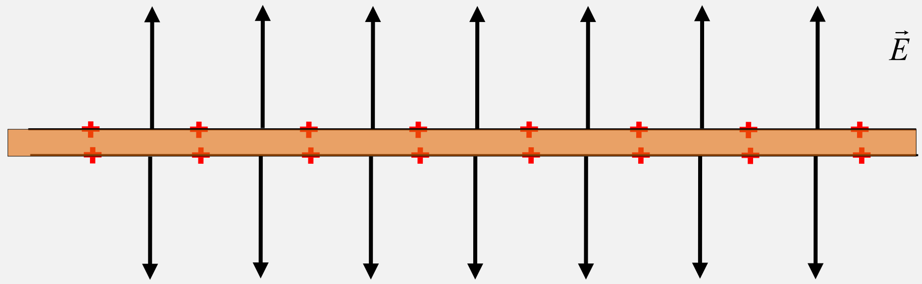
A much more common scenario is to have *two parallel sheets of charge*, one positive and one negative, having the same charge density.

Let us consider first the case where both sheets are insulators, the charge has been sprayed on. On bringing the two sheets close, the charges will be unable to move, and the electric fields from the two planes add, from the Principle of Superposition, giving:



Electric field from two uniform parallel planes of charge: one positive, one negative, with equal charge densities σ : the field $E = \sigma / \varepsilon_0$ between the planes, $E = 0$ outside the planes.

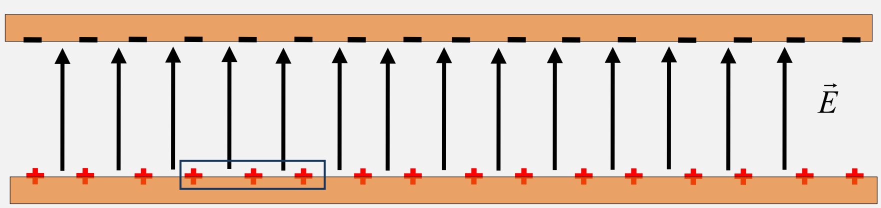
Actually, the sheets are usually conductors—in fact, almost all capacitors have this basic structure. To see how the charges move as the conducting sheets are brought close, we'll first look at the charge distribution on a single conducting sheet of finite thickness:



Electric field from a uniformly charged plane conductor of finite thickness: in the electrostatic situation, there is as always no field inside the conductor, the charges form equal layers on the two sides

Suppose we now take two such conducting planes with equal charge densities, but of opposite signs, and put them close and parallel. What happens?

The positive and negative charges will attract each other, and move to be as close together as possible. That is, all the charges will move to the inside surfaces of the conductors:



Electric field for two uniformly charged plane conductors of finite thickness: no field inside the conductors, or outside the two planes: the charges are moved to the inside surfaces by their mutual attraction.

Note that the charge density σ on the lower conductor's top surface generates a field of strength σ / ε_0. This can be understood by considering a pillbox Gaussian surface which encloses that top surface (see diagram): the Gaussian surface has field E through its top, but *no electric field* through its bottom, which is *inside* the conductor, where $E = 0$.

6 Gravitational and Electrostatic Potentials

The Gravitational Analogy

As we've discussed, the gravitational field from a point mass and the electrostatic field from a point charge both go down with distance as $1/r^2$, and both fields satisfy the Superposition Principle. It might seem at first glance that electric fields are just going to follow the patterns set by gravitational fields—but of course, there's one huge difference! Electric charges can attract or repel, but there's no gravitational repulsion between masses.

You might think antimatter would repel matter, but experimentally it doesn't—all kinds of matter attract gravitationally. Then there's the so-called Dark Energy in the universe, which apparently causes everything to fly apart, *but* is only important on a cosmic scale. And, don't confuse *that* with Dark *Matter*, an as yet unidentified kind of matter we know must be there because its gravitational *attraction* is clear from the orbiting rate of stars in rotating galaxies, but it also has little effect on anything much smaller than a galaxy.

Near-Earth Gravitational Potential *mgh* and Its Electrical Equivalent

Let's begin by reviewing the Earth's gravitational field in this room. We can take it to be uniformly downward: a mass m will feel a downward force mg, and doubling the mass doubles the force. That is, the gravitational force on a mass m is $\vec{F} = m\vec{g}$ where $\vec{g}$ is a downward pointing vector of length g, the gravitational field strength.

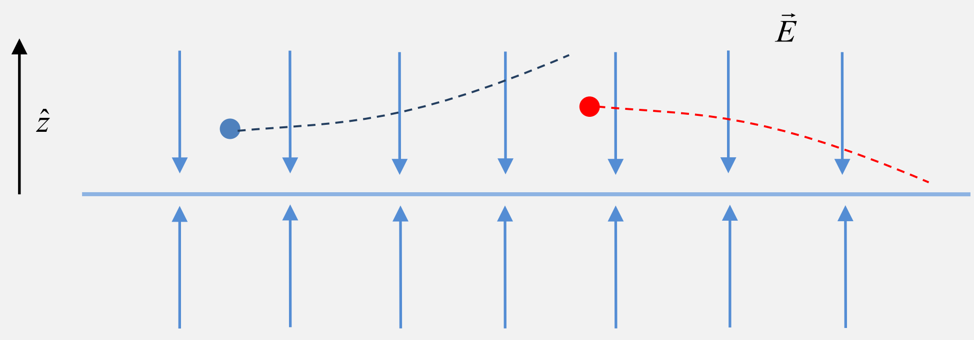
It takes work to lift a mass m from a point *A* to a higher point *B* against this gravitational pull: to be precise, as discussed earlier, it takes work $W = \int_A^B \left(-m\vec{g}\right)\cdot\overrightarrow{ds}$, where $\overrightarrow{ds}$ is an incremental step on the path, and to move the mass at a steady rate we need to exactly counteract the gravitational force, that is we must exert a force $-m\vec{g}$, so the work done for the step $\overrightarrow{ds} = -m\vec{g}\cdot\overrightarrow{ds}$. Since this is a dot product, the only displacement that requires work is that in the upward direction, and it is easy to see that the total work done against the gravitational force on raising a mass m from *A* to *B* is $W = mg\left(h_B - h_A\right)$, where h_B is the height of point *B* above the ground. This work is stored by the system—it can be recovered simply by allowing the mass to fall back: it accelerates and gains kinetic energy equal to the work needed to raise it in the first place. That's why it's called "potential energy".

This leads naturally to the definition of a *gravitational potential*

$$U\left(h\right) = gh,$$

so $mU\left(\vec{r}\right) = mgh$ is a measure of the potential energy stored by a mass m as a function of position. Following normal usage, we denote height by h rather than z. There is of course the usual ambiguity concerning what "ground level" we take as $h = 0$, but it is irrelevant in practice as we're always interested in potential energy *differences*.

The electrostatic analogy to gravity near the Earth's surface is the electric field in the region above an infinite, uniformly negatively charged insulating plane: *we covered this in lecture 3*.



Electric field from a uniform negative plane of charge, showing trajectories of charged particles.

-

The electric field has uniform strength and points towards the plane. The force on a charge *q* is $\vec{F} = q\vec{E}$.

Since there is no reason for this plane to be horizontal, we'll measure distance away from the plane as z, so $\hat{z}$ is a unit vector normal to the plane. By precise analogy with the gravitational discussion, the work needed to move a charge q along a path in this field is $W = \int_A^B \left(-q\vec{E}\right)\cdot\overrightarrow{ds}$, and, without further ado, we can define an electrostatic potential

$$V\left(z\right) = Ez = \left(\sigma / 2\varepsilon_0\right)z,$$

where σ is the charge density (this is for a charged plane, see diagram: remember that for a uniform layer of charge σ on the surface of a thick *conductor*, there will be *no factor two* in the denominator, because there is no field going into the conductor, all the field from the charge layer is on one side).

Now this was a negatively charged plane, so a positively charged particle projected upwards from this plane will follow a parabolic path and come back down, just as a mass will in the gravitational field in this room.

A *negatively* charged particle, on the other hand, will follow a parabolic path upwards! To see this, consider a particle projected parallel to the plane but some distance above it. Particles with the same mass but opposite charges will follow paths that are up-down mirror images of each other.

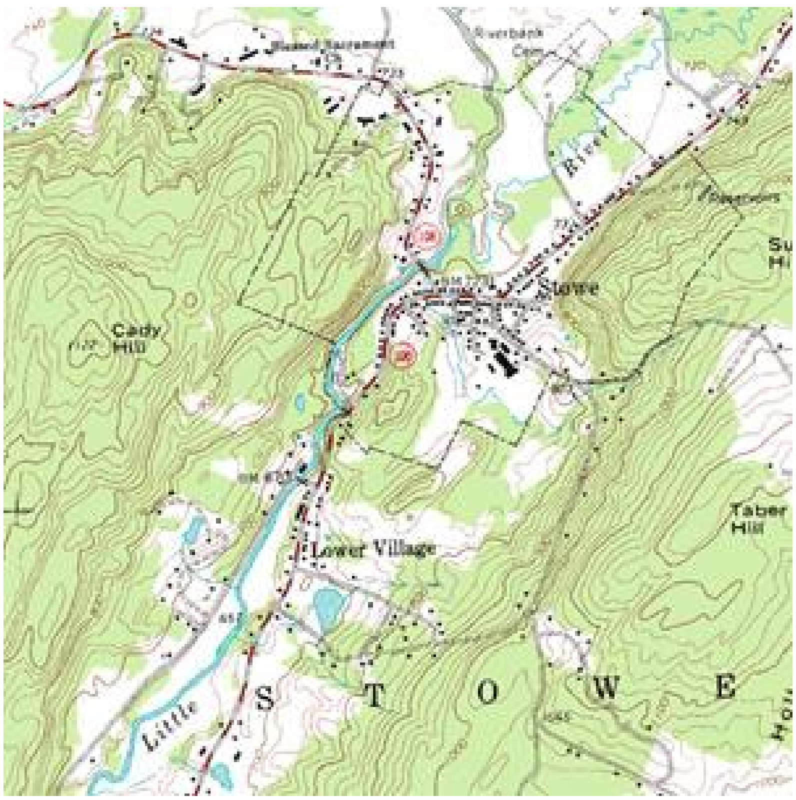
In practice, a uniform electric field as described above is well approximated in the space between two oppositely charged parallel planes. It is also a good approximation to the field near the surface of a charged conductor—near enough for the conductor to appear flat.

Gravitational Equipotentials are Contour Lines

Detailed maps of the countryside for hiking often include *contour lines* joining points at the same height h, in our language, points at the same gravitational potential. Walking along a contour line means you do no work against gravity. Of course, on level ground the force of gravity on you is balanced by the normal force from the ground, but if the ground is sloping and you walk uphill a distance $\Delta\vec{s}$ you do work against the component of gravity parallel to the ground, $-m\vec{g}\cdot\overrightarrow{\Delta s}=mg\Delta h.$ For a given speed, you work at the fastest rate (as of course you know!) by going straight uphill, meaning *perpendicular to the contour line*.

Notice from the map that the distance between contours is a measure of the steepness of the slope, an (inverse) measure of the strength of the gravitational field you are working against.

Simple example: for a map of a conical hill, the equipotentials would be concentric circles.



The Gravitational Analogy at Larger Distances

At distances comparable to the size of the Earth, the gravitational field has the familiar inverse square form $\vec{g}(\vec{r})=-GM_E\hat{r}/r^2$.

The work done, and therefore the potential energy difference, on a path in this field is (as discussed above)

$$W = \int_A^B \left(-m\vec{g}\left(\vec{r}\right)\right)\cdot\overrightarrow{ds},$$

except that the gravitational field is no longer constant.

As before, the dot product denotes that *work is only done when there is displacement in the direction of the force*: here this means *displacement in the radial direction*, directly outwards. So, if *A* is at $\vec{r}_A$ from the center of the Earth, and *B* at $\vec{r}_B$, the gravitational potential energy difference for a mass $m=1$ is

$$U\left(\vec{r}_B\right)-U\left(\vec{r}_A\right)=GM_E\int\limits_{\vec{r}_A}^{\vec{r}_B}\frac{\hat{r}\cdot\overrightarrow{ds}}{r^2}=GM_E\int\limits_{r_A}^{r_B}\frac{dr}{r^2}=GM_E\left(\frac{1}{r_B}-\frac{1}{r_A}\right).$$

The very reasonable convention is to take the zero of gravitational potential energy to be at infinity, because in calculating total potential energies, we don't want to have to take account of stars in the next galaxy. This means that, outside the Earth's surface, the gravitational potential energy from the Earth's field is

$$U\left(\vec{r}\right)=-\frac{GM_E}{r},\ \ r>r_E.$$

A mass resting at the Earth's surface has therefore a *negative* total energy (potential plus kinetic), a mass at rest far away has essentially zero total energy—so to get a mass away from the Earth it must be given a kinetic energy sufficient to get it up the potential hill: this corresponds to the escape velocity.

We are thinking in three dimensions and the equipotentials here are spherical *surfaces.*

Point Charges and Superposition

Switching now from gravity to the electrostatic analogy, the potential difference between two points in the field of a point charge Q (or outside a *spherically symmetric* charge distribution having total charge Q) is:

$$V_{\vec{r}_B}-V_{\vec{r}_A}=-\int\limits_{\vec{r}_A}^{\vec{r}_B}\vec{E}\cdot\overrightarrow{ds}=-\frac{Q}{4\pi\varepsilon_0}\int\limits_{\vec{r}_A}^{\vec{r}_B}\frac{\hat{r}\cdot\overrightarrow{ds}}{r^2}=-\frac{Q}{4\pi\varepsilon_0}\int\limits_{r_A}^{r_B}\frac{dr}{r^2}=\frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{r_B}-\frac{1}{r_A}\right).$$

In words, as with gravity, the potential difference is the work done against the field per unit charge/mass on moving from point *A* to point *B*.

For an actual *point* charge, assuming one could exist, it is clear that for small enough r the formula must break down (there cannot be infinite energies!) but even for electrons within atoms the formula is extremely accurate. (It does break down at electron scattering energies reached in particle accelerators: the field energy density becomes strong enough that from quantum mechanics virtual particle creation plays a significant role, this is termed *quantum electrodynamics*.)

Thus the potential in the electric field of a point charge is (taking it zero at infinity):

$$V(\vec{r}) = \frac{Q}{4\pi\varepsilon_0 r}$$

Notice the sign!

If a positive charge is released in the field of a fixed positive charge, it will shoot away, and have kinetic energy far away. This is the *opposite* of the "escape velocity" scenario—that applies for a negative charge in the field of a fixed positive charge.

The *Principle of Superposition works for potential energies* just as it does for electric fields, since the potential energy difference is the sum of contributions from the different fields in the integral, so

$$V(\vec{r}) = \frac{1}{4\pi\varepsilon_0}\left(\frac{Q_1}{r_1} + \frac{Q_2}{r_2} + \frac{Q_3}{r_3} + \cdots\right)$$

and for continuous charge distributions, the sum becomes an integral.

An Atomic Energy Unit: the Electron Volt

For everyday life, the joule is a convenient unit of charge—one amp is a charge flow of one coulomb per second. Similarly, the volt, one joule per coulomb, is a convenient potential energy unit. But when we're analyzing energy transfer at the molecular level, the natural unit of charge is the electron charge (or minus it).

7 Field Lines, Equipotentials and the Dipole

Getting the Electric Field from the Potential

Important! It's usually easier to compute the potential than the electric field for a given charge distribution, since, for the field, one must sum over vectors. A very simple example is finding the potential on the axis of a uniform ring of charge—for a point on the axis, all the charge is at the same distance so no integral is necessary. You can go on (see next paragraph) to easily find the electric field component pointing along the axis (but the perpendicular field is tougher!)

So, suppose we have the potential as a function of position. How do we use it to get the electric field at a particular point (x, y, z)?

Write the formula for potential difference between two points separated by an infinitesimal distance dx:

$$V(x+dx, y, z) - V(x, y, z) = -\int_{(x,y,z)}^{(x+dx,y,z)} \vec{E} \cdot \overrightarrow{ds} = -E_x dx$$

from which

$$E_x = -\frac{\partial V(x,y,z)}{\partial x}$$

where the special derivative symbol means *partial* differentiation: V is a function of three variables, but we're holding two of them constant—only allowing x to vary.

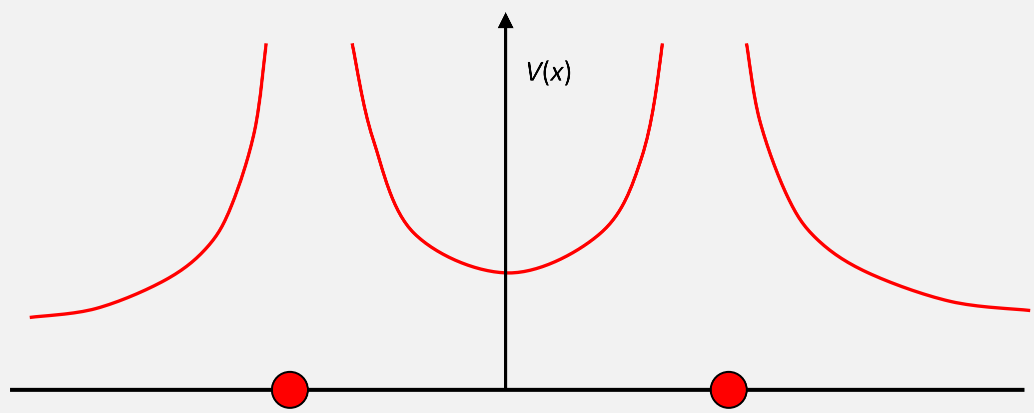
This formula, plus those in the other two directions, are often combined in a vector notation, written:

$$\vec{E} = -\vec{\nabla} V, \text{ or } \vec{E} = -\mathbf{grad} V.$$

In other words, the electric field in a particular direction is the negative of the slope of the potential in that direction: it's worth looking at a couple of examples to see this in action.

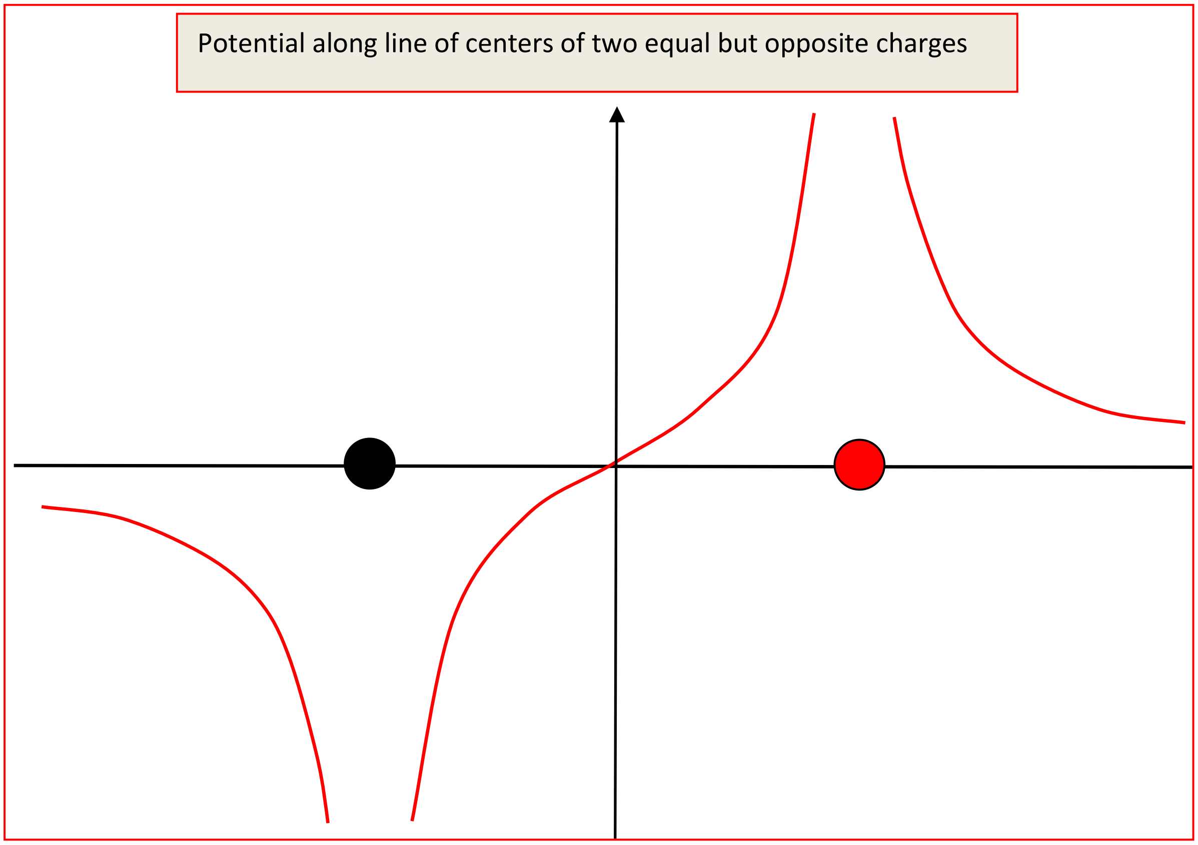
Potential for Two Charges

First, consider two equal positive charges, let's say on the x-axis at $+a$ and $-a$, and think about the electric potential and electric field on the axis. This keeps it simple: the electric field points along the axis. The potential plotted along the *x*-axis looks like:



Potential from two equal positive charges plotted on line through center of charges

Over to the far right, the potential is sloping downwards, so the $\vec{E}$ field is pointing in the positive x-direction. Exactly half way between the charges, the potential bottoms out, that is, its slope is zero: so the electric field is zero at that point—not surprising, since a small positive charge there will be repelled equally by the two positive charges. In fact, the electric field changes sign (it's minus the potential slope) on going through that point. Note as well, though, that the value of the potential at that low point is *not zero*: if we moved away from that point in the y-direction, we'd be going downhill. (Check that by finding the electric field direction at a point on the y-axis.) The second example is a positive charge at $+a$, a negative charge at $-a$ on the x-axis. Now the potential along the axis looks like this:

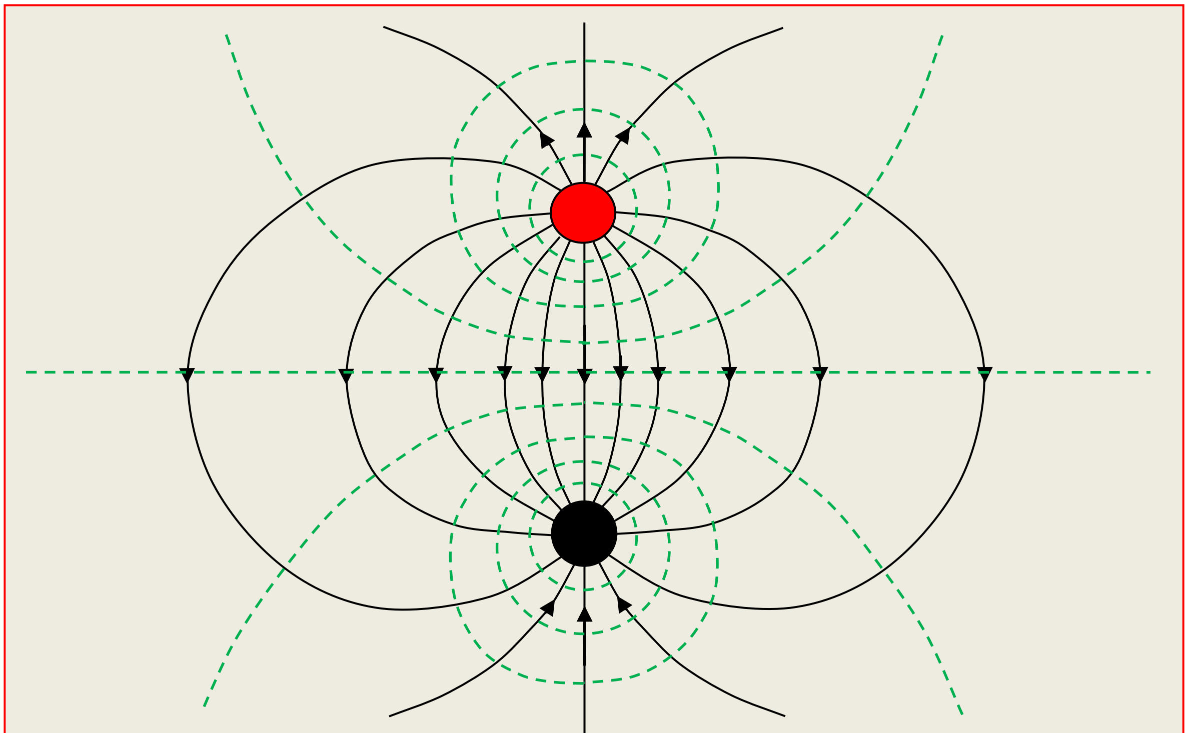


In this case, the electric field between the two charges is always strong and in the negative x-direction.

Revisiting the Dipole: Field Lines and Equipotentials

If you sprinkle iron filings in the field from a magnet, they line up along the field direction, and you can draw what Faraday and Maxwell called "lines of force", parallel to the field at each point, to construct a picture of the field. We'll call them field lines (a more usual term).

We drew the field lines in lecture 3, now we'll add the equipotentials. Recall from the discussion above that they intersect at right angles, so, for example, for a pair of opposite charges we find:



Electric field lines of force and **equipotentials** for a **dipole**: charges *Q*, -*Q* a fixed distance ℓ apart

Note: Field strength and line spacing: this is a two-dimensional cut through a three-dimensional field. We can see where the field is strongest because the field lines are more concentrated. But this is tricky! We can draw field lines where we want. Furthermore, consider the field from a single small charged ball. If we draw the lines coming from the ball's surface equally spaced, then In this two-dimensional representation, the spacing between adjacent lines increases with distance proportional to r, suggesting the field strength goes down as $1/r$. But that isn't right, the lines are really coming out of the ball in three dimensions, and their density, the measure of field strength, actually goes down as $1/r^2$. So these two-dimensional representations of three-dimensional fields can be useful, but be careful—they're not quantitatively correct.

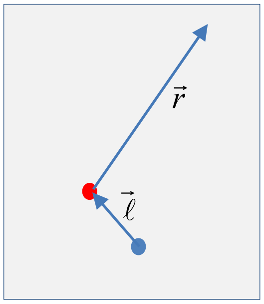
Dipole Moments and Potential

Non-ionized molecules are overall electrostatically neutral (no net charge) but can have dipole moments, meaning the center of the positive charge is not the center of the negative charge. As in the diagram above (and previously mentioned in lecture 3, where we discussed a dipole in an external field) we can represent the dipole moment as equal but opposite charges $\pm Q$ separated by a vector distance $\vec{\ell}$.

The *dipole moment* is written

$$\vec{p} = Q\vec{\ell}.$$

The potential from the two charges at a point $\vec{r}$ measured from the positive charge as origin is



$$V(\vec{r}) = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{|\vec{r}|} - \frac{1}{|\vec{r}+\vec{\ell}|}\right).$$

If we assume we are at a distance much greater than the size of the molecule $r \gg \ell$, we can approximate

$$|\vec{r}+\vec{\ell}| = \sqrt{r^2 + 2\vec{r}\cdot\vec{\ell} + \ell^2} \cong r\left(1 + \frac{\vec{r}\cdot\vec{\ell}}{r^2}\right),$$

and

$$\frac{1}{|\vec{r}+\vec{\ell}|} \cong \frac{1}{r}\left(1 - \frac{\vec{r}\cdot\vec{\ell}}{r^2}\right) = \frac{1}{r} - \frac{\vec{r}\cdot\vec{\ell}}{r^3}.$$

Putting these together, the potential at a point $\vec{r}$ from a dipole of strength $p = Q\vec{\ell}$ with $r \gg \ell$ is

$$V(\vec{r}) = \frac{Q}{4\pi\varepsilon_0}\left(\frac{1}{|\vec{r}|} - \frac{1}{|\vec{r}+\vec{\ell}|}\right) \cong \frac{Q}{4\pi\varepsilon_0}\frac{\vec{r}\cdot\vec{\ell}}{r^3} = \frac{1}{4\pi\varepsilon_0}\frac{p\cos\theta}{r^2}.$$

Notice the dipole electric *field* strength, $\vec{E} = -\vec{\nabla}V$, decreases for $r \gg \ell$ as $1/r^3$.

For large distances, if an *ionized* molecule has total charge Q_{tot} the field strength very far away has dominant contribution Q_{tot}/r^2, this is sometimes referred to as the *monopole* field to distinguish it from the dipole field.

For general compact charge distributions, there are more terms, starting with the *quadrupole*, down by another factor of $1/r$: think for example of charge $2Q$ at the origin and two charges $-Q$ displaced equally from the origin in opposite directions.

8 Capacitance

Introduction: Charging a Sphere; Definition of Capacitance

A capacitor is a device for holding electrical charge. Of course, any electrically isolated macroscopic object can hold some charge, but the term *capacitor* is only used for conductors, so the whole object is raised to the same potential when the charge is added.

Perhaps the simplest example of a capacitor is a conducting sphere of radius R. As we found earlier, a charge Q on the sphere generates an electrical field outside the sphere of magnitude $E = (1/4\pi\varepsilon_0)(Q/r^2)$, so the potential at the surface of the sphere $V = (1/4\pi\varepsilon_0)Q/R = Q/C$ with

$$C = 4\pi\varepsilon_0 R.$$

That is, the charge Q of the sphere is linearly proportional to the voltage V, and the coefficient $Q/V = C$ is termed the capacitance.

In our system of units, the charge is measured in coulombs, and the capacitance which is raised in potential by one volt if one coulomb of charge is added is called a one farad capacitor, in honor of Michael Faraday. This is a pretty big sphere: recall $1/4\pi\varepsilon_0 = 9\times 10^9$, so if $C = 4\pi\varepsilon_0 R = 1$, we have $R = 9\times 10^9\,\mathrm{m}$, more than ten times the radius of the Sun!

If we need to store significant quantities of charge, spheres are not the best way to go (although a sphere *is* used in the van der Graaff machine).

Parallel Plates

Far more common are capacitors made of parallel plates of conductors: in the simplest case, two flat plates of area A are placed parallel a distance d apart, where d is much smaller than the linear size of the plates. This configuration was discussed in detail in lecture 5, so we'll just take the results from there. We take it that d is sufficiently small that the field between the plates is uniform, and the field outside the plates from the charge on the plates is negligible.

When connected to a battery, one plate to the positive and one to the negative terminal, charge flows on to the plates in equal (but of course opposite sign) amounts: if charge Q flows to the positive plate, it has charge density $\sigma = Q/A$, giving a uniform electric field outwards from each side $E = \sigma/2\varepsilon_0 = Q/2A\varepsilon_0$. This is the field from the positive sheet only, the field between the sheets has an equal contribution from the negative sheet, so

$$E = Q/A\varepsilon_0.$$

The voltage difference between the plates is then

$$V = Ed = Qd/A\varepsilon_0.$$

It follows immediately from the definition of capacitance, $V = Q/C$, that

$$C = \varepsilon_0 A/d$$

for the parallel plate capacitor.

Capacitance of a Coaxial Cable

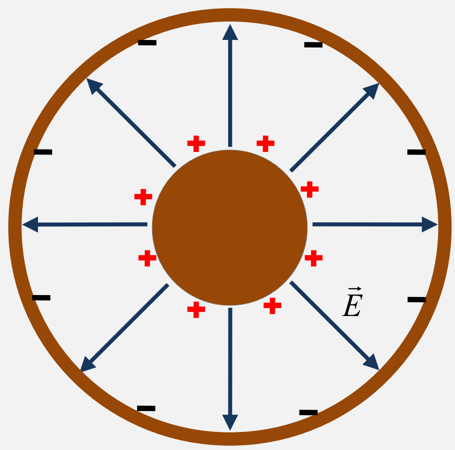
Recall from lecture 5 (where this diagram appears) the field configuration in a coaxial cable: the electric field strength between the inner solid copper wire and the outer encasing copper cylinder is given by $E(r) = \lambda / 2\pi r \varepsilon_0$, from Gauss' Law, where λ is the charge per meter on the wire (and the cylinder, of course). The voltage difference between the two cylinders is therefore, from a simple integration

$$V = \int_{r_1}^{r_2} E(r)\,dr = (\lambda / 2\pi\varepsilon_0)\ln(r_2 / r_1)$$

so the capacitance of a length ℓ is

$$C = Q/V = \ell\lambda / (2\pi\varepsilon_0)\ln(r_2 / r_1)$$
$$= 2\pi\varepsilon_0\ell / \ln(r_2 / r_1).$$

As we shall see later, this is important in analyzing the transmission of electromagnetic waves in coaxial cables—and that's the way the signal gets to your TV.



Typical electric field configuration in a coaxial cable, usually a copper cylinder and a central copper wire. The charge is on the **outside** surface of the **inner** conductor, the inside surface of the outer conductor.

Capacitors Big and Small

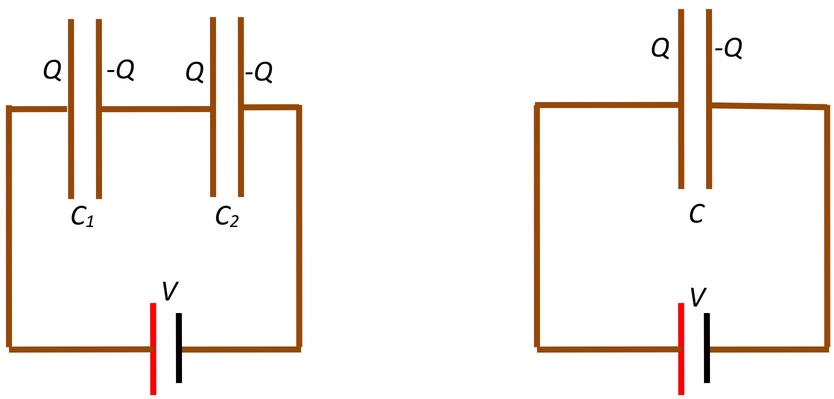
With parallel plates, we don't need a capacitor bigger than the Sun to get one farad. But it still has to be pretty big, if we keep the gap between plates an easily visible size, say 0.1mm. The reason is that ε_0 is so small (8.85×10^{-12}). The area has to be of order square kilometers! Traditional commercial capacitors lessen the gap by having plates separated by a thin layer of insulator (which is also a dielectric—see later) and roll up the plates into a many layered roll. Still, it's difficult to get much above millifarads this way in a compact capacitor.

A real breakthrough came some years ago with the realization that aluminum oxidizes almost immediately on exposure to air, that the oxide layer that forms is about a micron (10^{-6} meters) thick, and is a good insulator. Capacitors were then made by putting conducting paste on to oxidized aluminum. The paste was one plate, the aluminum metal the other. More recently, capacitors have been manufactured with a layer of insulator a few atoms thick. This is another factor of 1,000 down in thickness. At the same time, the area has been vastly increased by using activated carbon, a solid which is actually many tiny granules pressed close, but with most of their surface still exposed, to give hundreds of square meters of surface in an ordinary size jar (your lungs have a similar structure—and comparable surface area, necessary to absorb oxygen at the required rate).

The only drawback is that the insulating layer cannot resist more than three volts or so, this being the typical voltage to excite an atom. However, these new capacitors are measured in kilofarads, and will soon be competitive with conventional batteries in hybrid cars. One advantage over batteries is the rapidity with which capacitors can absorb and deliver power.

At the other end of the scale, dynamic rapid access memory (DRAM) in computers stores information in millions of capacitors of microscopic size, arranges in rows and columns on a chip. These are measured in femtofarads (10^{-15} farads). So capacitors are currently being manufactured over a range of sizes 10^{18}!

Combining Capacitors in Circuits: Series and Parallel



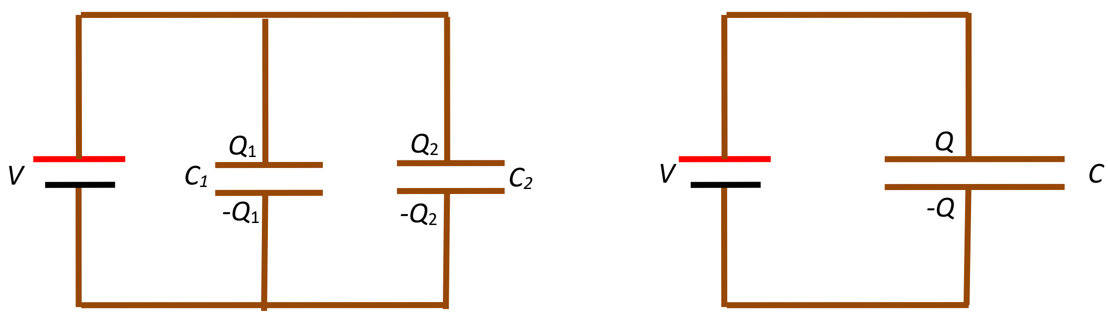
Two capacitors C_1, C_2 in **series** and the equivalent single capacitor C. The charges on C_1, C_2 must be the same, since the charge on the plate of C_1 not connected to the battery must come from C_2. Since the single capacitor C behaves identically, for given V it also draws charge Q.

Two capacitors that appear one after the other in a circuit, as shown above, are said to be in *series*. They can be replaced by a single capacitor which will behave identically, meaning if the two were in a black box with just the wires coming out the side, by testing with various voltages and noting the charge flowing in, you wouldn't be able to tell. But, given C_1, C_2 what is the value of the equivalent capacitor C? The key is to note that if the $C_1 + C_2$ combination is subject to the same external voltage as the single C, the same charge must flow in—otherwise, the C wouldn't be equivalent. Also, equally important, in the combination the Q's on the two capacitors must be the same, since the Q from the battery on C_1 will draw $-Q$ from C_2 as shown.

Now consider the total voltage drop on going around the circuits. For C, it's $V = Q/C$. For $C_1 + C_2$, there is voltage drop across each capacitor, so the total $V = Q/C_1 + Q/C_2$. These voltage drops for the two circuits are equal, so for *capacitances in series,*

$$\frac{1}{C}=\frac{1}{C_1}+\frac{1}{C_2},$$
$$C=\frac{C_1C_2}{C_1+C_2}\quad \text{(series)}.$$

For *capacitances in parallel,* at given voltage V, the total charge drawn from the battery by the two capacitors, Q_1+Q_2 must equal the charge Q drawn by the single equivalent capacitor, from which



Two capacitors C_1, C_2 in *parallel* and the equivalent single capacitor C. The **voltages** across C_1, C_2, C are all V. Since the single capacitor C and the parallel pair C_1, C_2 behave identically, the total charge Q_1+Q_2 drawn from the battery must equal Q.

$$C=\frac{Q}{V}=\frac{Q_1}{V}+\frac{Q_2}{V},$$
$$C=C_1+C_2\quad \text{(parallel)}.$$

Simple Picture of Adding Two Capacitors

Suppose we take two capacitors which are physically parallel metal plates: the capacitances are $C_1=\varepsilon_0 A_1/d_1, C_2=\varepsilon_0 A_2/d_2$. First, take two for which $d_1=d_2$. Place them side by side, and connect the two top plates, then the two bottom plates: put them in parallel. Obviously, the combined capacitor C has the same $d=d_1=d_2$, and $A=A_1+A_2$, so $C=C_1+C_2$. Next, take two having $A_1=A_2$ and put them in series: For the combined C, $A=A_1=A_2$, $d=d_1+d_2$, the result follows.

9 Energy in Capacitors

Work Done in Charging a Capacitor

Suppose we put a charge Q on a capacitance C, thereby raising its potential to $V = Q/C$. Obviously this takes work: as soon as there is any charge on the capacitor, it will repel further charge we put on, so we need to work against that electrostatic repulsion.

To be precise, when the capacitor has charge q it is at potential q/C, and bringing in from far away an incremental additional charge dq requires work equal to the potential energy that small extra charge has gained, that is, $dW = qdq/C$.

The total energy stored in the capacitance once it has charge Q is equal to the total work needed to get the charge there, that is,

$$U = \int_0^Q \frac{qdq}{C} = \frac{Q^2}{2C} = \tfrac{1}{2}QV,$$

using $V = Q/C$.

This Energy is Stored in the Electric Field!

To show this claim makes sense, we'll consider a few examples, starting with the parallel plate capacitor. Suppose as usual we have uniformly charged (σ coulombs/sq m) plates of area A (so $Q = A\sigma$) separated by a distance d which is much smaller than the linear dimensions of the plates, so we will have a constant electric field inside (meaning between the plates), and a negligible field from the charged plates outside. Then the energy

$$U = \tfrac{1}{2}QV = \tfrac{1}{2}A\sigma Ed = \tfrac{1}{2}\varepsilon_0 E^2 Ad = \tfrac{1}{2}\varepsilon_0 E^2 \times \text{volume}$$

where we used $V = Ed$, $\sigma = \varepsilon_0 E$.

This turns out to be always true: an electrostatic field is a store of energy, with energy density $\tfrac{1}{2}\varepsilon_0 E^2$ *per unit volume.*

Field Energy for a Charged Sphere

To see that the idea of storing energy in an electric field is not just about parallel plates, consider a spherical conductor of radius R carrying a charge Q. The sphere is then at potential $V = Q/4\pi\varepsilon_0 R$, so the energy stored is $U = \tfrac{1}{2}QV = Q^2/8\pi\varepsilon_0 R$.

We can derive this value as the total electric field energy density: since it's a conductor, there is zero field for $r < R$, so the total electric field energy

$$U = \int_R^\infty \tfrac{1}{2}\varepsilon_0 E^2 dv = \tfrac{1}{2}\varepsilon_0 \int_R^\infty \left(\frac{Q}{4\pi\varepsilon_0 r^2}\right)^2 4\pi r^2 dr = \frac{Q^2}{8\pi\varepsilon_0 R}.$$

(Here dv is the volume increment.)

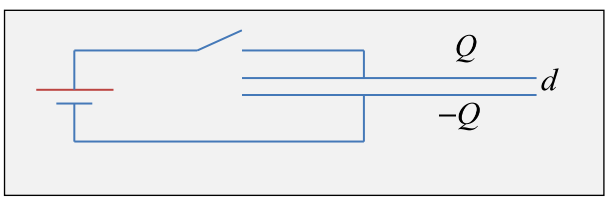
*Energy Stored in the Field for Two Charged Spheres

If we have two such spheres, one positive and one negative, *far apart*, the total energy in the electric field is just twice that for a single sphere. If they're on top of each other, there is no field at all! (one could be a thin skin on the other.) If we draw them distance d apart, $d \gg R$, their own fields dominate out to a distance of d or so from each sphere center, beyond which they approximately cancel, leaving the much weaker dipole field. Compared with the potential energy for the spheres being infinitely far apart, then, it's as if they have each lost the field energy which was outside a sphere of radius d or so, (which would be the *total* field energy of a charged sphere of radius d,) and since there are two of them, the field energy lost is about $Q^2 / 4\pi\varepsilon_0 d$. This, then, is the energy that must be supplied to get their separation from d to infinity.

Of course, the above is a very hand waving argument, and not to be trusted within factors of 2, etc.—but the answer does happen to be right.

Pulling a Disconnected Charged Capacitor Apart

Thinking in terms of energy stored in the electric field gives some insight into the force needed to pull capacitor plates apart. Suppose we pull the plates from separation d to $2d$.



Assume first that the capacitor is charged but disconnected, so the charge Q stays the same. Then the field strength $E = \sigma / \varepsilon_0 = Q / A\varepsilon_0$ remains constant, but there is now an extra volume Ad of field— new field—the field energy storage has doubled.

Question: The increase in energy is $\tfrac{1}{2}QV = \tfrac{1}{2}QEd$. But, assuming we held one plate fixed and moved the other, which has charge Q, through distance d, and the electric field between the plates is E, where does the ½ come from?

Answer: The field between the plates, remember, is a superposition of equal fields from the two plates. When we move one plate, it doesn't do work against its own field, which moves with it.

Suppose now both plates were equally *positively* charged: work can be extracted from this system by the plates pushing each other apart. What's going on with the fields now? In this case, there is constant field *outside* the plates, no field in between. As they move apart, the zero-field region expands at the expense of the field region, so field disappears. That's where the work done by the pushing plates comes from.

Pulling a *Connected* Charged Capacitor Apart

Suppose now we pull apart the plates of a charged capacitor *keeping it connected to the battery*, so V doesn't change. How does *this* affect the electric field? Since $V = Ed$, if we go from d to $2d$, the field strength drops by a factor of 2, since the voltage V is constant, but the *volume* of field doubles. Since the field energy is proportional to E^2, not E, the total energy stored in the field volume is down by a factor of 2. This should be no surprise: remember we can write $U = \frac{1}{2}CV^2$, and doubling d halves the capacitance, so at constant V it halves the energy stored.

But it *did* take work to pull the plates apart at constant V: where did that energy go? The answer is: into the battery! On halving the capacitance at constant voltage, we must have lost half the original charge Q. This $\frac{1}{2}Q$ goes into the battery against the voltage V, so the battery is recharged with restored energy $\frac{1}{2}QV$.

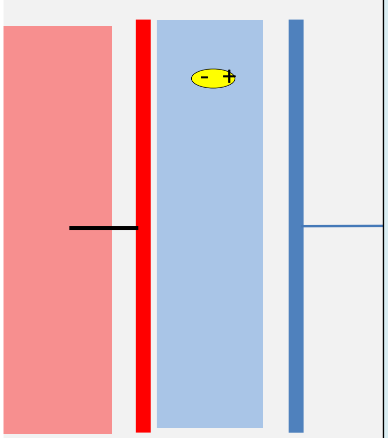
But only half that energy pumped into the battery came from energy stored in the capacitor's electric field: the rest came from work done dragging the plates apart. Let's check that: if the plates have separation x, the field strength $E = V/x$, the field from a single plate is half that, $V/2x$, and the charge on the plates is proportional to E. Therefore, since the initial force on one plate was $QE/2 = Q^2/2\varepsilon_0 A$, and this was at separation d, so at an increased distance x it is down to $\left(Q^2/2\varepsilon_0 A\right)\left(d^2/x^2\right)$, and the work needed to double the plate separation is

$$\Delta U = \frac{Q^2}{2\varepsilon_0 A}\int_d^{2d}\frac{d^2}{x^2}dx = \frac{Q^2 d}{4\varepsilon_0 A} = \frac{Q^2}{4C}.$$

The bottom line is: the work done pulling the plates apart, plus the energy consequently lost from the capacitor, both go into recharging the battery—no energy has disappeared.

Dielectrics

Typically, capacitors have some material between the plates, if only to keep them apart and prevent electrical shorting. This way the plates can be placed closer together, decreasing d and thereby increasing C. But it turns out that the right material increases C in another way as well: all materials respond to some extent to an electric field. Remember the water molecule is an electrical dipole, so in an electric field it will tend to line up with its positive $-+$ end closer to the negatively charged plate, in a parallel plate capacitor. But even molecules that



A slab of dielectric between parallel charged plates. The dielectric molecules distort so that negative charge moves slightly towards the positive outside plate, equivalent to adding an extra layer of negative charge on that surface, and a positive layer on the right.

are not dipoles will become dipoles to some degree in a field, since all molecules have positive nuclei and negative electrons, and the different charges will be displaced in opposite directions by the field. If we think of a neutral solid as a block of positive charge on top of a block of negative charge, this has the effect of moving the blocks in opposite directions, with the result that charge cancellation still holds in the bulk, but there are layers of negative charge and positive charge respectively on the sides next to the positive and negative plates.

These charges are of course still bound inside their molecules, so cannot get away—but they have the effect of partially cancelling the charges on the plates Q, so the electric field is lessened, therefore the voltage corresponding to given charge on the plates is less, so the capacitance is enhanced.

Relative Permittivity

The relative permittivity K (a.k.a. the *dielectric constant*) is the factor for a given material by which the capacitance increases on introducing the material to fill the gap between the plates: $C = KC_0$. In other words, it's a measure of how much a material *permits* an applied magnetic field to store energy in it, relative to the vacuum.

As you might guess, it's very high for water, 80, (the molecular dipole moments align, at least for a steady field) but water is not a great choice for a dielectric, for obvious reasons. For many ordinary materials (paper, oil, glass) K is around 4, well worth using.

Note that the charges within the dielectric move towards the opposite charges on the capacitor plates, so pulling the dielectric out will take work against this attraction. For a charged but disconnected capacitor, this work goes into building up the larger electric field between the plates corresponding to the lower capacitance but same Q when the dielectric is gone. If the dielectric is removed with the capacitor connected to a battery, the work goes into charging the battery.

Capacitors for Energy Storage

There have been breakthrough in capacitor design in recent years, so now kilofarad capacitors are available. However, currently (2024) storage capability is 5 – 10% that of a lithium ion battery per kilogram, although current design developments using carbon nanotubes could substantially increase that.

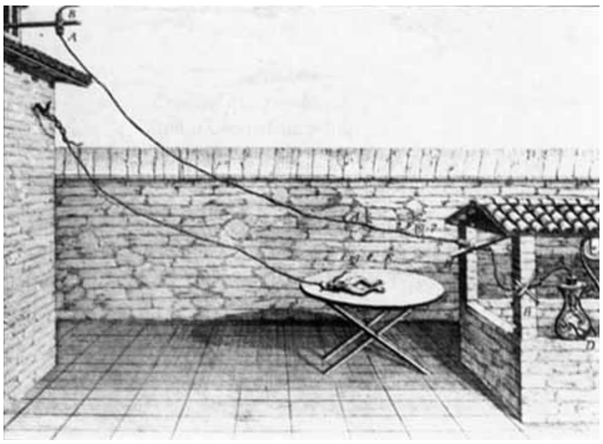
There are some short-hop bus systems using capacitors, https://en.wikipedia.org/wiki/Capacitor_electric_vehicle, the low storage capacity is good for a mile or two, and charging is almost instant.

The rapidity of discharge is also useful in situations where an explosive burst of power is needed.

10 Electric Charge and Current

Electricity and Frog's Legs

- In 1771, Luigi Galvani, at the University of Bologna, was dissecting frog's legs at a table that also had an electrostatic generator. He found by accident that the legs twitched in response to a charge, and were far more sensitive than the best electroscopes. He tried to detect atmospheric electricity.
    - He found instead that electricity was generated by touching the legs with *dissimilar metals*.



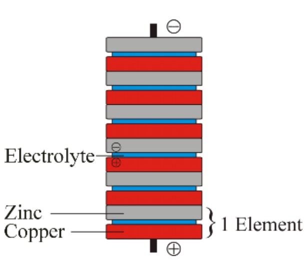
Reviving Dead Criminals?

Galvani's nephew, Giovanni Aldini, a showman, electrified corpses just after decapitation at a prison in London, with various muscular reactions.

This was the inspiration behind Frankenstein.

It also led to the belief that electricity was the "life force", the essential non-material component of living matter, absent in ordinary inanimate matter. This idea was demolished by Volta.

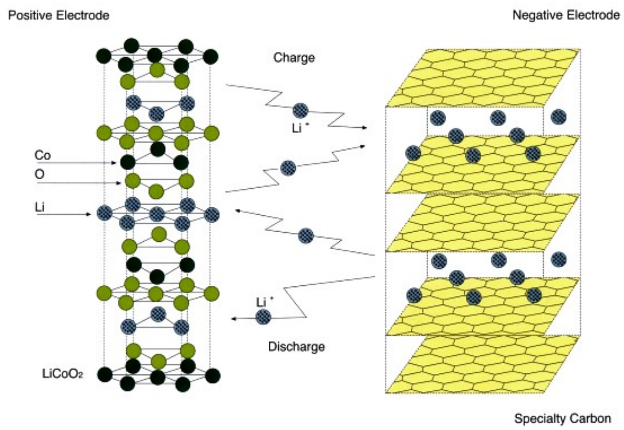
Volta's Pile



Galvani's colleague Volta was the first to realize that using different metals to touch the frog's leg was crucial to producing electricity, and in fact the leg could be replaced with *cardboard soaked in brine*: no sign of life!

He built a pile of such metal pairs—the first such battery—with dubious medical applications, as diagrammed here.

A Modern Battery: Lithium Ion



Lithium ions Li^+ are *very tiny*: remember H, He, Li, …they are He atoms with an extra nuclear charge. They can fit between atomic layers in graphite, to which they bond, but bond more strongly in $LiCoO_2$. Charging is by attracting them from the $LiCoO_2$ into the graphite by pumping in electrons.

Batteries, Circuits, Currents

The two terminals of a battery, called *electrodes*, are immersed in an *electrolyte*. Positive ions are formed at one electrode by atoms depositing electrons.

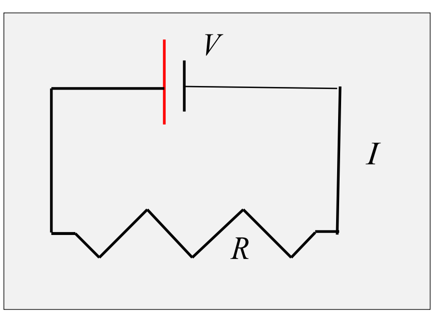
For suitably chosen materials, energy is generated by these electrons *flowing round an outside wire* to take part in a chemical reaction (or just rejoin the ions) at the other electrode.

The "outside wire" is the circuit. Flow is measured in coulombs per sec, called Amperes.

Ohm's Law

Ohm found experimentally in 1825 that for a given piece of wire, the current, labeled I, was directly proportional to the applied voltage (from number of battery cells) V, and wrote it as $I = V/R$, where V is in volts, I in amps.

R is called the *resistance* of the wire, and is *measured in ohms*: one volt sends one amp through one ohm.



These are the standard symbols for a battery and a resistance: remember the standard "current" is really electrons flowing the other way!

Electric and Water Currents Compared

It's sometimes useful to think of electric current down a wire as resembling water flowing down a pipe.

Pressure difference between two ends of a water pipe corresponds to *voltage* difference between the ends of a wire.

Flow rate is determined by pressure gradient: a water pipe twice as long drops twice the pressure during flow, in electrical terms, *a wire twice as long has twice the resistance*.

Resistance and Cross-Section Area

Suppose we take two identical wires, having the same area of cross section A, and twist them together to make one wire.

When this is done, it's found (not surprising) that the combination delivers twice the current of a single wire for the same voltage.

But effectively we've doubled the cross-section area: so R is proportional to $1/A$.

(In fact, this turns out not to be true for real fluid, like water, in a pipe, where doubling the cross-section area for the same shape more than doubles the flow rate—but it's accurate for electric current flow.)

Resistance and Resistivity

To summarize: for a given material (say, copper) the resistance of a piece of uniform wire is proportional to its length ℓ and inversely proportional to its cross-sectional area A.

This is written:

$$R = \frac{\rho \ell}{A},$$

where ρ is the *resistivity*. For copper, $\rho = 1.68 \times 10^{-8}\ \Omega \cdot \text{m}$.

Electric Power

Remember voltage is a measure of potential energy of electric charge, and if one coulomb drops through a potential difference of one volt it loses one joule of potential energy.

So a current of I amps flowing through a wire with V volts potential difference between the ends is losing IV joules per sec.

This energy appears as heat in the wire: the electric field accelerates the electrons, which then bump into impurities, vibrations, and defects in the wire, and are slowed down to begin accelerating again, like a sloping pinball machine.

Power and Energy Usage

Using Ohm's law, we can write the power use of a resistive heater (or equivalent device, such as a bulb) in different ways:

$$P = IV = I^2 R = V^2 / R.$$

The unit is *watts*, meaning *joules per second*.

Electric meters measure *total energy* usage: adding up how much power is drawn for how long, the standard unit is the kilowatt hour:

1 kWh = 1,000x3,600J = 3.6MJ.

Energy Storage

A more recent energy unit is the watt.hour, used in energy storage capability: a tesla battery is currently (2024) capable of storing around 270 w.h/kg, so a four-kilogram battery is needed to store 1 kWh. Tesla is now building batteries to store wind power, single units store 3.9 MWh, weigh forty tons, cost $1.4M. Efficiency is about 93%.

For comparison, in Bath County, Virginia, an energy storage facility simply pumps water into a high reservoir, then releases it to drive turbines as it returns to a lower reservoir. When running, it generates 3000 MW and can run for eight hours. The overall efficiency is 79%, cost about $1.6B in 1985.

11 Microscopic Theory of Electric Current

Ohm's Law and Drude Theory

Ohm's Law, written down in 1825, relates the current I through a resistance R to the applied voltage V,

$$I = V / R.$$

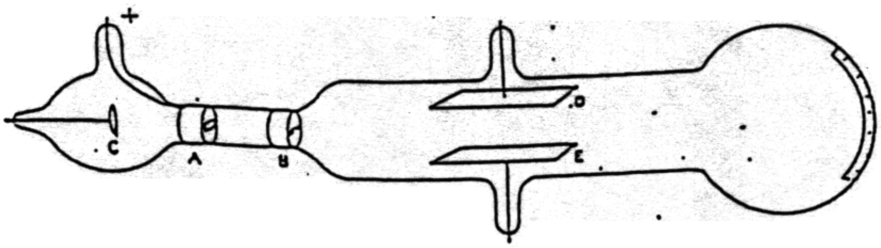
He wrote this after extensive experimentation, finding it to be true for many metal resistances over wide temperature ranges (although R itself was usually temperature dependent).

Naturally, a simple law of such wide validity was a tempting target for theoretical speculation. Yet the first serious attempt to explain it with a model came seventy-five year later! Why did it take so long?

In fact, over this long period, no one had any idea what the "electric fluid" constituting the current looked like on a microscopic scale. In Ohm's Austria, for example, like other German speaking states, almost no one believed in the existence of atoms—many imagined solids to be continuous at all scales.

One exception was Boltzmann (in the 1890's) who had read Maxwell's work on the kinetic theory of gases, and extended it, but had difficulty convincing his colleagues, and was told that on publishing he would not be allowed to mention "atoms". This negative reception probably contributed to his suicide.

The necessary conceptual breakthrough in understanding matter came seventy years later, in England, when J. J. Thomson (in 1897) discovered the *electron*. He was analyzing the rays in a cathode ray tube, here is his own diagram:



The "tube" is the overall glass enclosure, with a good vacuum inside. The cathode is marked C on the far left. In the experiment, the cathode is raised to a high negative voltage, the metal rings A and B (see figure) are positive. At high enough voltage difference, rays were seen to emanate from C, some passed through slits in A and B, and struck the right-hand end of the apparatus, where the glass glowed. The effect is much enhanced by coating the glass at that end with a phosphor which shines brightly when struck by the rays.

Next, (see figure), Thomson charged plates B and C oppositely so the rays passed through a sideways electric field. The rays were deflected! *They must be charged particles*. As we'll soon discuss, measuring the deflection for given electric field strength, and also measuring deflection by a magnetic field, a simple calculation yielded the ratio charge/mass. Assuming the charge equaled that of the hydrogen ion in magnitude (as was soon established), their mass was 2,000 times smaller than the lightest atom.

These must be the electric fluid! And, they must have been inside atoms in the cathode. Thomson repeated the experiment with different materials making up the cathode. He always found the same particles emitted. Evidently these "electrons" were present in all atoms.

Thomson then suggested the *plum pudding* model of the atom: a sphere of positive charge with electrons embedded like raisins in a pudding. Electrons could be knocked out, leaving a positively charged ion—the electron had the right charge value for this to work.

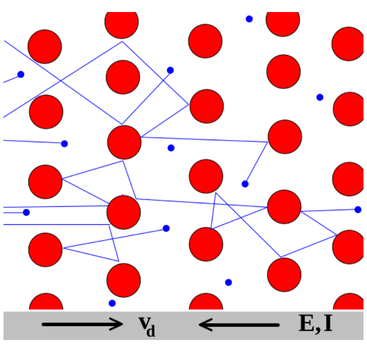
A crystalline material like copper was known to have a regular array of atoms, and it was known that some solids, like salt, NaCl, form by the atoms becoming ionized and electrostatic attraction binds them into a cubic grid, where the nearest neighbors of an ion all have the opposite sign charge. In the pudding picture, this could be understood by saying an electron leaves each Na and becomes part of a neighboring Cl.

But Cu must be a little different from NaCl. For one thing, it conducts electricity. This suggests that as the atoms bind to form the solid, some fraction of the electrons remain unattached, free to move through the crystal—and these become a current when an electric field is applied.

In 1894, Paul Drude became a top professor at the University of Leipzig, having gained a Ph. D. studying the diffraction of light by crystals, most naturally understood in terms of scattering from a regular array of atoms or ions. This was also the time when Maxwell's theories were becoming more widely accepted in Germany, both his electromagnetic equations and his kinetic theory of gases.

So by the late 1890's *for the first time* concepts were available to formulate a semi-plausible theory of electrical conduction: Maxwell's theory of gases (extended by Boltzmann) could be applied to the "gas" of electrons, including Maxwell's predicted temperature-dependent velocity distribution. Of course, Maxwell's theory was for gas molecules in an otherwise empty box. Hopefully the velocity distribution would still more or less work for the electrons in a "box" already containing the rows of ions, even though the electrons must keep colliding with the ions? Also, the electrons repelled each other, but, at least on average, this was compensated by the background attraction from the ions.

Anyway, despite these obvious objections, Drude hypothesized that, statistically, the electron gas had the symmetric Maxwell velocity distribution when in zero external electric field, although each individual electron had a probability dt/τ of scattering in time dt, and would scatter to some other random velocity in the Maxwell distribution, with no memory of its velocity just before scattering.



Now switch on an electric field $\vec{E}$ at $t = 0$. Each electron will accelerate to gain velocity $\vec{v} = e\vec{E}t/m$ (added to its original Maxwell distribution velocity) until it collides with an ion, at which point the process will repeat.

Suppose at some later time t we take a freeze frame shot of the electrons: with the given scattering rate, the average time interval since the last scattering is τ, so the *average* electron velocity (called the *drift velocity*) is

$$\vec{v}_{\mathrm{d}} = e\vec{E}\tau/m.$$

(Remember the pre-collision velocities average to zero.)

Taking the density of electrons to be n, the electric current density

$$\vec{j} = ne\vec{v}_{\mathrm{d}} = \frac{ne^2\tau}{m}\vec{E}.$$

Defining the conductivity

$$\sigma = \frac{ne^2\tau}{m},$$

we have Ohm's law in the form

$$\vec{j} = \sigma\vec{E},$$

relating the current density to the electric field strength.

As we discussed in the last lecture, to recover the traditional form $I = V / R$ for current down a wire, take the wire to have cross-section area A and length ℓ, so the total current $I = jA$, assuming it's uniform across the area, which it is, and the voltage drop down the piece of wire $V = E\ell$ so

$$j = I / A = \sigma V / \ell,$$

and

$$R = V / I = \ell / A\sigma.$$

Standard notation is to define the *resistivity*

$$\rho = \frac{1}{\sigma},$$

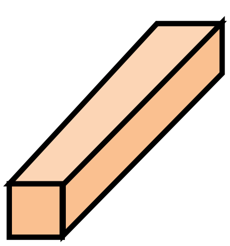
so the resistance R of a length ℓ of wire having cross-section A is

$$R = \frac{\rho\ell}{A}.$$

Checking Drude's Model Against Experiment

First, it does predict Ohm's law. But to get some picture of how it relates to reality, it would be useful to find, for example, the scattering time τ.

By 1900, the charge e and mass m of the electron were known, as was the electron density n, at least approximately, so the drift velocity could be measured as follows:



Take a piece of copper wire, say 1mmx1mm cross section, 1m long carrying 5 amps.

This is 1cc of Cu, about 10 gms, about 10^{23} conduction electrons (assuming one per atom), about 15,000 Coulombs of electron charge.

Therefore, at 5 amps (C/sec) it takes 3000 secs for an electron to drift 1m.

Bottom line: the drift velocity is of order 0.0003 m/sec.

This wire has resistance $R = \rho\ell / A \approx 0.02\Omega$ so from Ohm's law $E \approx 0.1\,\mathrm{V/m}$.
This field will accelerate the electrons, $ma = eE$, approximate acceleration = 2x10^{10} m/s ^{2} This reaches the drift velocity in about 0.5x10 $^{-14}$ seconds, that must be the time τ.

So how far does the electron move on average between collisions? (This is the *mean free path*, often labeled ℓ.)

Drude assumed the electrons had a Maxwell velocity distribution, which would give an average velocity $\overline{v}$ at temperature T:

$$\tfrac{1}{2}m\overline{v}^2 = \tfrac{3}{2}k_\mathrm{B}T,$$

with k_B Boltzmann's constant. This gives an average electron speed of order 10^5 m/sec, so the distance between collisions, the mean free path, is of order 0.5x10^{-9} m, close to the interionic distance, a reasonable sounding result.

Other insights from Drude's theory included the behavior of a current when a magnetic field was added (the Hall effect), and the previously mysterious relationship between electrical conductivity and heat conductivity (in metals heat is mainly conducted by electrons).

What About Quantum Mechanics?

However, it turned out that although the picture of an electron gas with random scatterers was essentially correct, the advent of quantum mechanics changed everything. An electron in a regular crystal is wavelike, and passes through a perfect crystal at zero temperature without scattering, like light through glass. It *is* scattered by impurities, and by thermal vibrations. This explains why resistivity of metals increases approximately linearly with temperature over a wide range. The formula is:

$$\rho_\mathrm{T} = \rho_0\left[1+\alpha\left(T-T_0\right)\right].$$

An old incandescent (not LED) bulb has a tungsten wire at about 3300K, and α = 0.0045, from which the resistivity is not far off being proportional to absolute temperature.

Experimentally (and theoretically) the electron mean free path is at least an order of magnitude more than Drude's model suggests, yet the mean free *time*, which we found from the drift velocity, must still be the same. Going back to a particle viewpoint, this means the electrons are going at least an order of magnitude faster than Maxwell's velocity distribution predicts. Turns out they are not at all like a classical gas. First, the wavelike nature means that, as in the Bohr atom, only certain wavelength are allowed (so a whole number of wave oscillations fit in the box), and, second, there's the exclusion principle: no two electrons can be in the same quantum state. This means the free electrons are forced up into high velocity states—this is the essential point. Unfortunately, a fuller explanation would take several lectures— meaning a proper introduction to quantum mechanics.

AC and DC

We'll discuss AC in much more detail later.

Batteries provide direct current, DC: it always flows in the same direction.

Almost all electric generators produce a voltage of sine wave form:

$$V = V_0 \sin 2\pi ft = V_0 \sin \omega t.$$

In a resistance R this drives an alternating current, AC,

$$I = \frac{V_0 \sin \omega t}{R} = I_0 \sin \omega t$$

and power

$$P = VI = I^2 R = I_0^2 R \sin^2 \omega t = \left(V_0^2 / R\right) \sin^2 \omega t.$$

So the power is rapidly oscillating, what matters in practice is almost always the average power.

The average value $\overline{\sin^2 \omega t} = \frac{1}{2}$ (from $\overline{\sin^2 \omega t} + \overline{\cos^2 \omega t} = 1$.)

We define the *root mean square* voltage V_{rms} by

$$V_{\mathrm{rms}} = \sqrt{\overline{V^2}} = V_0 / \sqrt{2},$$

so the average power

$$\overline{P} = V_{\mathrm{rms}}^2 / R.$$

The standard 120 V AC power is $V_{\mathrm{rms}} = 120V$, so the maximum voltage $V_0 = 120\sqrt{2} \approx 170$ V.

Semiconductors

In the Bohr model of the hydrogen atom, an electron circles around a proton.

An n-type semiconductor is a dielectric insulator which has been doped—atoms having one more electron than the insulator atoms are scattered into it.

The extra electron circles the dopant atom, but is loosely bound because the dielectric shields the electric field—it looks like a big Bohr atom. As the temperature is raised, these electrons break away from their atoms, and become available to conduct electricity.

Bottom Line: Conductivity *increases* with temperature.

Superconductors

A superconductor has exactly zero resistivity.

In 1911, mercury was discovered to superconduct (R = 0) when cooled below 4K.

Superconducting magnets are widely used, in MRI machines, etc.

There are now materials superconducting above the boiling point of liquid nitrogen, making long distance transmission lines feasible.

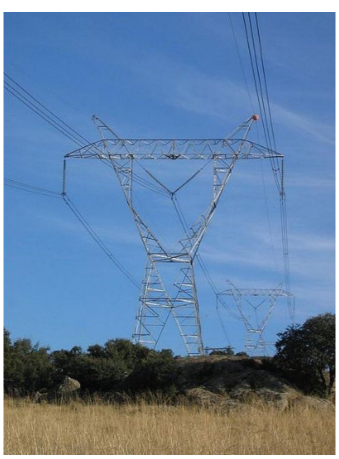
Superconductivity is a quantum phenomenon.

Why Bother with AC? (*we'll discuss this more later*)

Because, as we'll see, it's very easy to transform from high voltage to low voltage using transformers.

This means for long distance transmission we can use very high voltage, hence small currents and thinner wires, but transform to less dangerous low voltages for local use.

Long distance lines use aluminum wires. Copper is a better conductor, but is much heavier and more expensive. Steel is sometimes added for strength.



Sometimes DC is used for a single long line.

- This 3 gigawatt DC line (enough for 2 to 3 million households) transmits hydropower from the Columbia river to Los Angeles.
- At these distances, it gets tricky synchronizing the phase of AC power.

12 DC Circuits I

Introduction: Electromotive Force and Terminal Voltage

In this lecture, we'll analyze current flow in a network of resistances and include the possibility of batteries in some branches. We only address steady current flow, so do not include capacitances or inductances—these will be dealt with a little later.

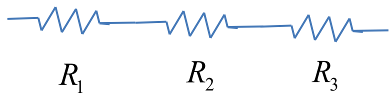
Beginning with the simplest case of a single battery, the potential difference that drives current originates in the chemical reactions inside the battery, at the surface of contact of the electrolyte and the terminals, called the anode and cathode. The two chemical reactions (releasing an electron at the anode to go around the circuit to combine chemically at the cathode) add to give a driving potential called the electromotive force, denoted by $\mathcal{E}$. This drives the current around the circuit but also through the battery itself, which has its own resistance, usually denoted by r. Thus the potential delivered outside, called the terminal voltage and denoted by V, is given by

$$V = \mathcal{E} - Ir.$$

Often r is small enough for this correction to be ignored.

Remark: don't worry too much about the names anode and cathode. Check the Wikipedia article. For one thing, the names are switched on recharging. Also, in vacuum tubes the heated element is always called the cathode. Just concentrate on how the electrons/ions are moving.

Resistances in Series and Parallel

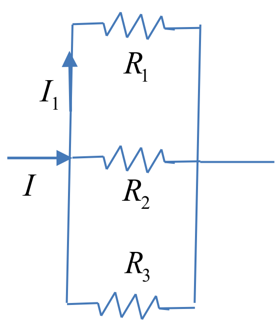


$$R = R_1 + R_2 + R_3$$

Applying Ohm's Law $V = IR$, the same current passes through all three resistances, so there are successive voltage drops IR_1, IR_2, IR_3 for a total voltage drop

$$V = IR_1 + IR_2 + IR_3 = IR$$

where $R = R_1 + R_2 + R_3$. Resistances in series just add.



$$I = I_1 + I_2 + I_3 = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

$$I = \frac{V}{R}$$

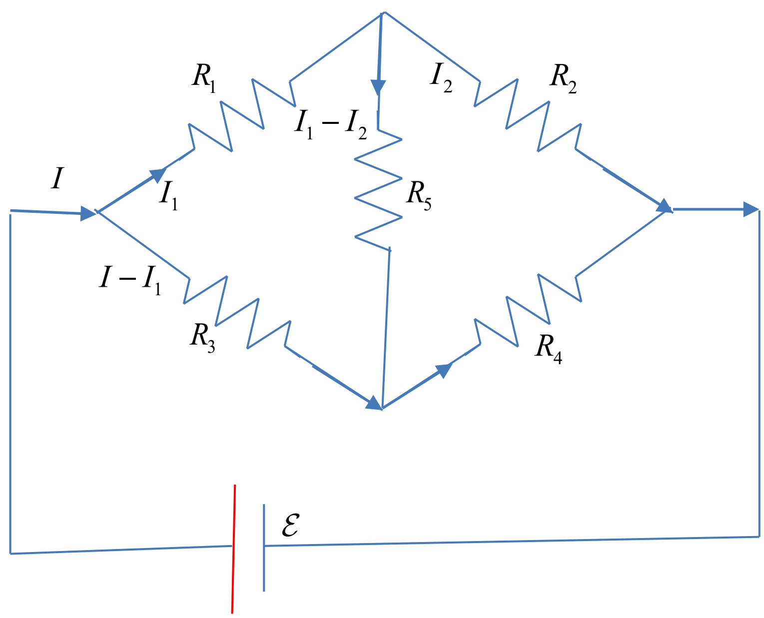
Parallel resistances all have the same voltage drop, the total resistance (see figure) is given by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}.$$

This is more obvious thinking in terms of the conductance (the inverse of the resistance): conductances just add, like parallel pipes conveying water.

More General Networks: Kirchhoff's Laws

We first consider a network of connected elements, as in this diagram, the individual elements can be resistances or batteries. (We'll add capacitances and inductances later.)



To analyze such a network, we label each element with its resistance R_i, the current I_i and the emf of any battery in that element $\mathcal{E}_i$. Then we use Kirchhoff's laws.

Kirchhoff Law #1: Junction Rule. At any connection point between elements, the total ingoing current must be zero.

In other words, charge cannot be piling up—the junction has no capacitance.

We have already applied this rule in the above diagram to reduce the number of unknown currents from five to two. The currents must be labeled with a value I_j and an arrow indicating direction.

Kirchhoff Law #2: Loop Rule. The potential drop across an element is $I_j R_j$. (plus possible battery term).

The total potential change on going round a closed loop back to the same point must be zero. The electric field is conservative, so

$$\sum_{\text{loop}} I_j R_j = 0.$$

If you take a walk on a hillside and finish at the same spot you began from, your total change in gravitational potential is zero. This is the same thing.

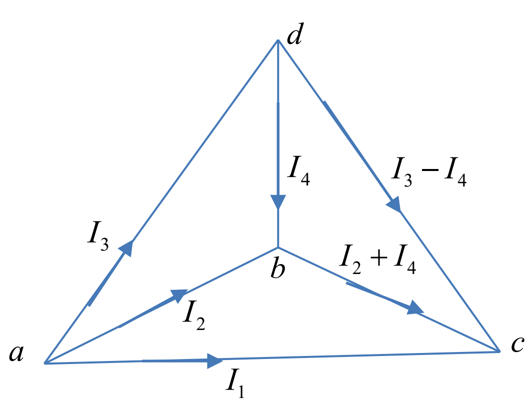
General Strategy for Solving Resistance Networks:

First, notice if there are resistances in series that can just be added, or in parallel that can be combined. (There may not be any.)

Second, label the current through each resistance, taking full advantage of the junction rule to minimize the number of unknowns.

Third, apply the loop rule to generate a number of equations equal to the number of unknown currents.

Solve these simultaneous linear equations to find the currents. You can then use Ohm's law to find the voltage drop for any resistance.



An Example: Compute the resistance R_{ac} of this network from a to c given that all lines are one ohm resistors except dc, which has resistance r.

In drawing the diagram, we've already applied the Junction Law at b, d to avoid introducing yet more unknown currents. This should always be done.

The total current flowing from a to c is

$I = I_1 + I_2 + I_3.$

Call the resistance of the network from a to c R_{ac}, then $V_{ac} = IR_{ac} = I_1$, the last being the voltage drop across the one ohm resistor ac.

So $R_{ac} = I_1 / I.$

Now we add to zero the voltage changes on going around loops, using $V = IR$ for each element, with $R = 1$ except for dc where $R = r.$

We have four unknown currents, but already have one equation above, given the external current I, so we need three loop equations. Here they are:

Loop abd : $I_2 = I_3 + I_4$. Loop abc : $I_1 = 2I_2 + I_4$. Loop acd : $I_1 = (1+r)I_3 - rI_4$.

From the first two, $I_1 = 2I_3 + 3I_4 = (1+r)I_3 - rI_4$, so $(r-1)I_3 = (3+r)I_4$.

It is now straightforward to express all currents as multiples of I_4 (do it!) to find

$$R_{ac} = \frac{I_1}{I_1 + I_2 + I_3} = \frac{3+5r}{8(1+r)}.$$

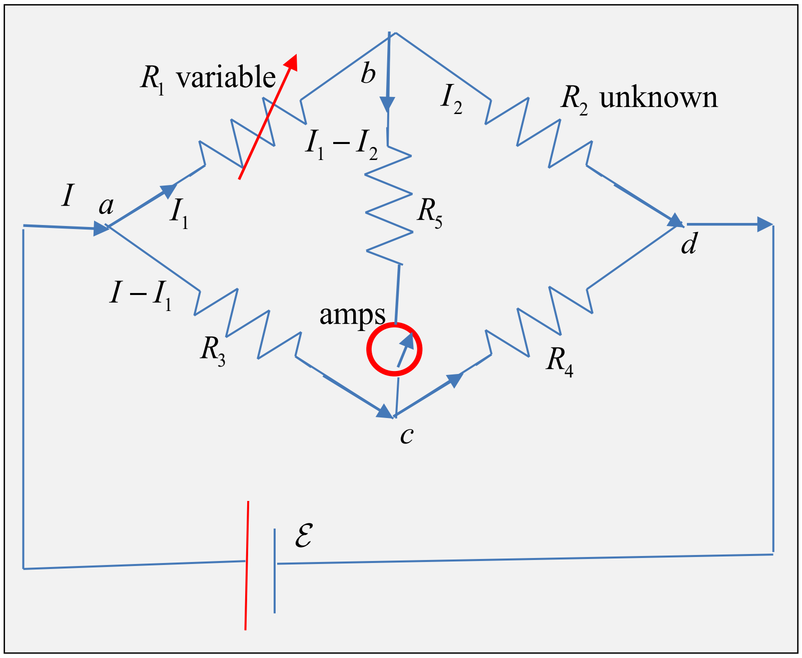
Exercise: Consider the three special cases $r = 0, 1, \infty$. See if you can find an easy way to find R_{ac} for each of these three cases, without going through all the work above.

13 DC Circuits II

Wheatstone Bridge

This is a circuit to measure an unknown resistance.

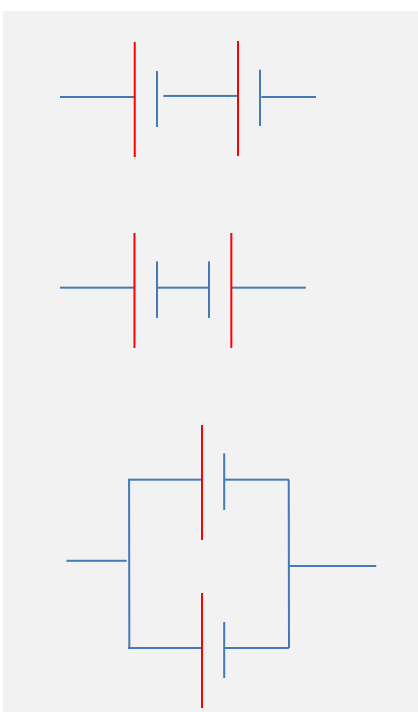
When the battery is connected, and the ammeter on the central vertical line registers zero current, the voltage drop ab must equal the voltage drop ac. Of course, the total voltage drop $abd = acd$ (it's a loop), from which, applying Ohm's law $V = IR$ to each resistance



$$\frac{R_1}{R_1+R_2}=\frac{R_3}{R_3+R_4}, \text{ so } \frac{R_1}{R_2}=\frac{R_3}{R_4}.$$

The resistances R_3, R_4 are fixed and known. The calibrated R_1 is varied until the central current reads zero, and the formula above then gives the unknown resistance R_2.

Series and Parallel EMFs

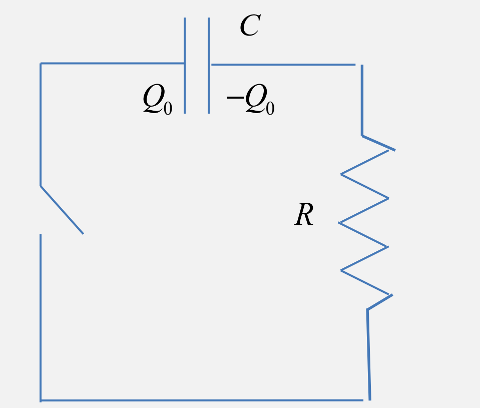


For batteries in series, the EMFs just add, as do the battery internal resistances. If equally matched batteries have opposite polarity in a completed circuit, nothing happens. If they're not equal, the stronger will drive the weaker backwards—in other words, charge it, reversing the internal chemical reactions. For a lithium ion battery this is especially easy to understand, the ions simply move from the $LiCoO_2$ to the graphite. In a hybrid car, this is happening almost all the time in stop-go driving, most braking just reverses the current, greatly improving gas mileage.

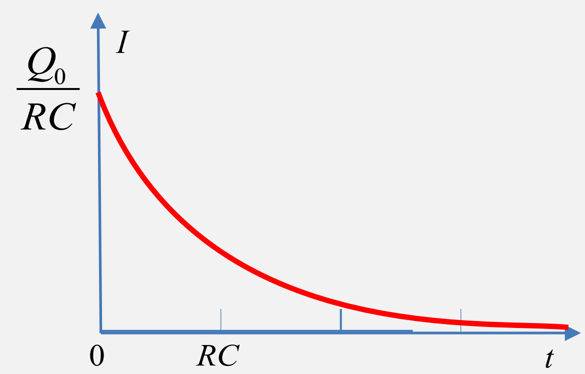
Identical batteries in parallel deliver the same EMF as a single battery, but with lower effective internal resistance, and, of course, longer life.

RC Circuits

Suppose we take a charged capacitor, charges $\pm Q_0$ on the two plates, then connect the plates using a wire having resistance R. The capacitor will discharge through the resistance, but how quickly? (Note that we are assuming negligible inductance in this circuit.)



Closing switch discharges capacitor



Current while discharging

Writing the charge on the capacitor as a function of time, $Q(t)$, the voltage across the capacitor $V(t) = Q(t)/C$ so from Ohm's law

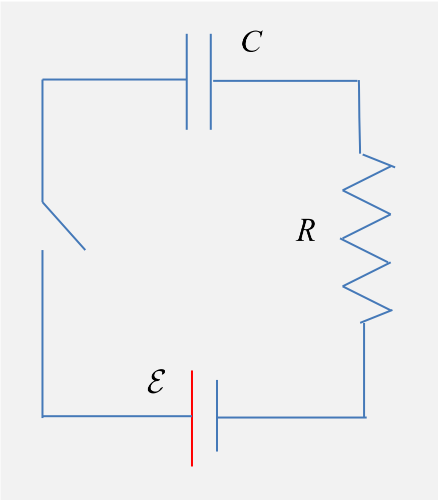
$$\frac{Q(t)}{C} = I(t)R = -\frac{dQ(t)}{dt}R,$$

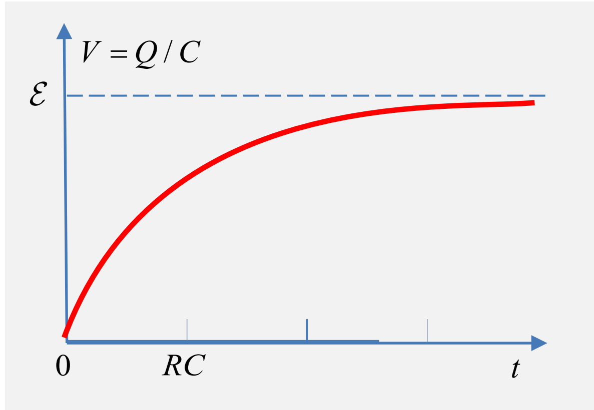
that is,

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

with solution $Q = Q_0 e^{-t/RC}$, and current $I = \frac{dQ}{dt} = \frac{Q_0}{RC} e^{-t/RC}$.

Charging a Capacitor





Charging a Capacitor

The capacitor is initially uncharged, when the switch is closed

$$\mathcal{E} = IR + Q/C = (dQ/dt)R + Q/C,$$

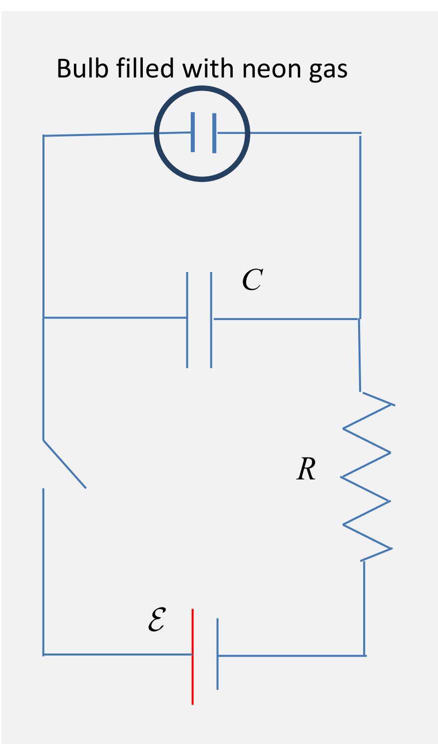
so

$$\frac{dQ}{dt} = -\frac{Q}{RC} + \frac{\mathcal{E}}{R},$$

and

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

Flashing Light



One application of an RC circuit is the kind of flashing light used at construction sites.

A gas filled bulb (typically neon) having two electrodes inside is connected between the two plates of the capacitor. On closing the switch, a large initial current flows into the capacitor, building charge—and therefore field—opposing the current. At low voltages, the neon gas is an insulator, but when the potential difference reaches a certain voltage, typically around 80V, the gas ionizes and a large current flows through the bulb with a flash of light, thereby discharging the capacitor to begin the cycle again.

14 Magnetism I

Early Observations and First Use

No doubt people were aware of magnetic and electrical phenomena much earlier, but the first recorded magnetic observations are from about 500 BC, the ancient Greeks. Rocks were observed to attract each other, and stick to iron nails in boots, in a place called Magnesia.

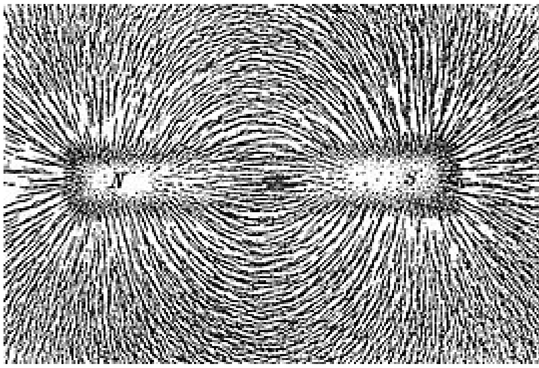
The first important use of magnetism was that of the compass in navigation. This was in approximately !000 AD in Northern Europe, and about the same time in China. (Apparently the Chinese had magnetic pointing devices earlier, but these were not used for navigation, only for finding most harmonious direction arrangements for furniture, etc., *feng shui*). The approximately simultaneous arrival of the navigational compass in Northern Europe and China suggests a common source, perhaps the Mongols, this is much discussed on the web, but the lack of documentation from this period renders it inconclusive, at least as far as I can see.

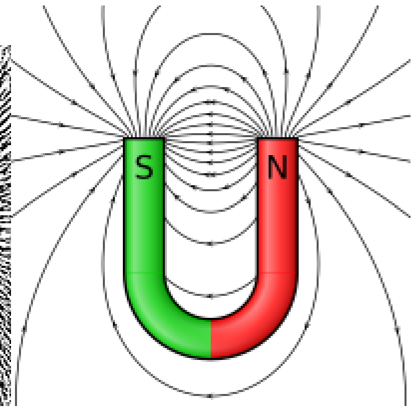
The first attempt to analyze magnetism from a recognizably modern point of view was the publication of *De Magnete* in 1600 by William Gilbert of St John's College, Cambridge (my college). He constructed a miniature earth (terrella) of lodestone and moved a small compass around its surface to demonstrate that the Earth itself was in fact a magnet. His work impressed Galileo, in fact their approaches were very similar, both had little patience for "authorities" who didn't do experiments.

For a fuller account of the development of these ideas, check out my notes here.

Magnets 101

Everyone is familiar with bar magnets, horseshoe magnets, and revealing the magnetic field by sprinkling iron filings, which line up with the field. There are always two "poles", labeled N and S from compass notation (N for "north seeking"). Like poles repel, unlikes attract.

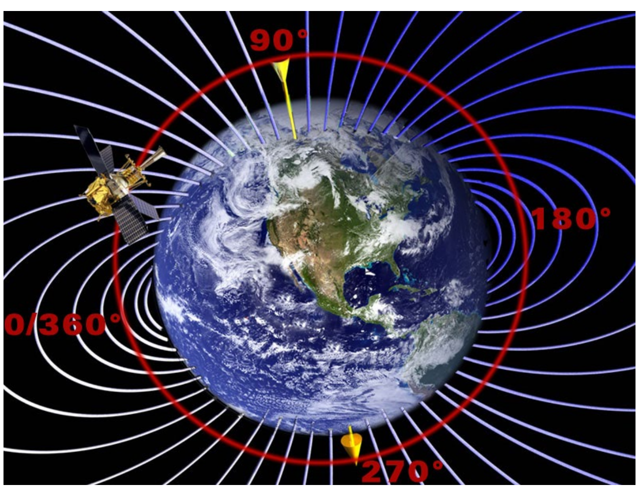




The iron filings make clear that the field pattern, especially some distance away, resembles the electric field from equal positive and negative charges close to each other, a dipole. But it's really quite different! You can have an isolated positive charge, you can't have an isolated magnetic north pole. If you break the bar in two, each piece will have its own N and S pole area.

The horseshoe magnet is a convenient configuration to concentrate the strong field into a small volume.

Powerful compact magnets are essential for building electric cars, and it turns out iron is not magnetic enough—the solution is to alloy with rare earths, in particular neodymium and dysprosium. Currently (2024) these are almost all mined in China, but can be found elsewhere, for example Ukraine and Greenland.

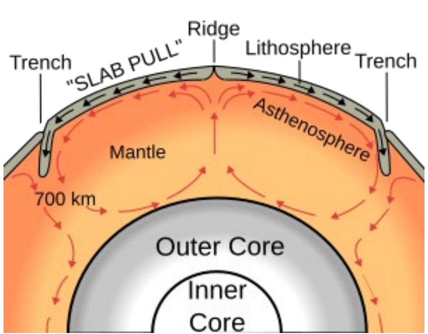


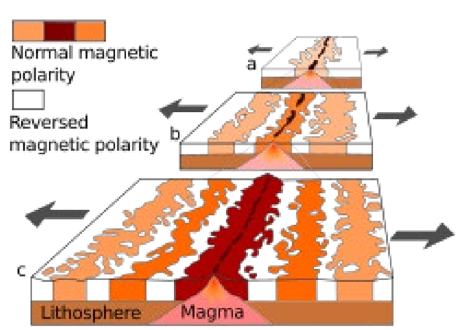
The Earth's Magnetic Field

Although the Earth's core is mainly iron, it is too hot to be magnetized (thermal vibrations knock the magnetic atoms out of line with each other).

The Earth's magnetic field is actually generated by electric currents in the outer core, driven by a combination of convection fluid currents and Corioli's forces. It is not a simple process.

The general shape is as perceived by Gilbert, a dipole (with the S end under the North pole, approximately). It is not in line with the Earth's axis of rotation, and in fact it has been proved that the fluid dynamics generating the dipole field would not work if it *was* in line.

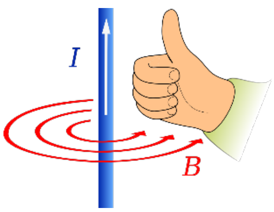




Seabed Stripes

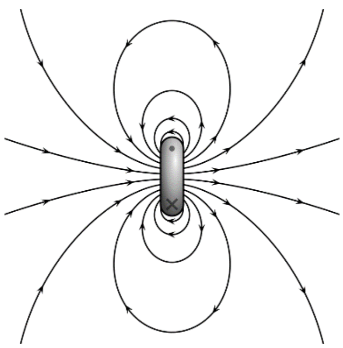
n the cold war (1950's) to better detect submarines magnetically, a detailed map of seabed magnetization in the Atlantic was made. It revealed a pattern of stripes of *reversed* magnetization, symmetric about the midatlantic ridge. This cast light on continental drift: hot materials well up at the ridge, get magnetized as they cool in the Earth's field, spread out both ways. And, it turns out, the Earth's magnetic field sometimes *reverses*, about every 300,000 years. Of course, any theory that explains the Earth's magnetization will have to include this.

Oersted's Great Discovery



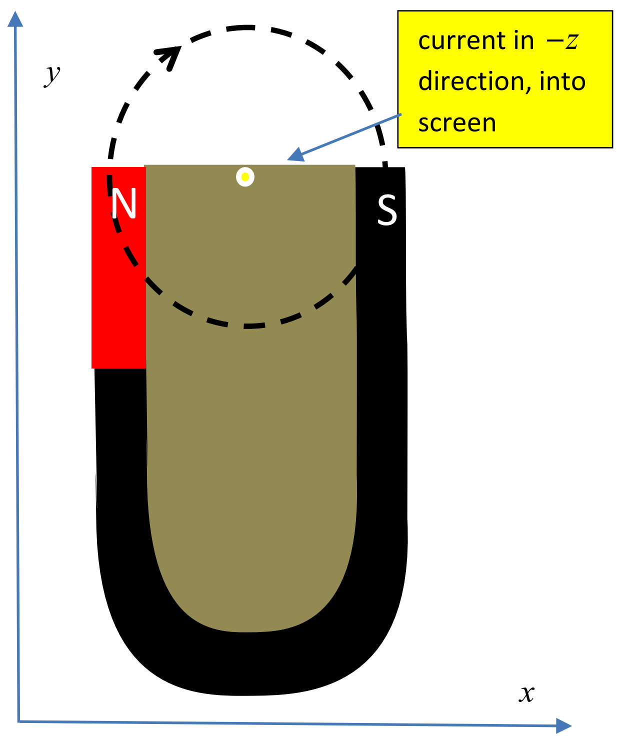
In 1820, the Danish physicist Oersted was the first to show electricity and magnetism were connected, by detecting the magnetic field of an electric current: remarkably, the field *circled around*, direction given by the right-hand rule (see figure). Here's a demo.

Notice how very different these field lines are from any possible static electric field. If you have a north pole, you can take it around a circle and end up at higher energy—evidently, unlike the static electric field, this field doesn't come from a simple potential.



Currents in Loops and Solenoids

Bending the wire into a circle, we can figure out the general shape of the field. Note that a *solenoid* (a series of connected loops) has a field resembling a bar magnet—but now we can see inside, and there are no poles. The magnetic field lines don't stop anywhere. No-one has ever detected a magnetic monopole, despite many expensive attempts, and at least one false alarm.



Force on a Horseshoe Magnet from Current in Wire and *Vice Versa*

In the diagram, the circle is a line of magnetic force from current going *downwards* in a wire perpendicular to the picture, passing through the yellow dot half way between the poles. Note the axes: this is a 3D problem.

The crucial point is that the magnetic force on the S pole is equal to and *parallel with* that on the N pole—the forces *add*!

So the horseshoe feels an upwards force, meaning in the y-direction in the diagram.

From Newton's Third Law, then, the wire must experience a downward ($-y$) force from the horseshoe's magnetic field. That field, of course, is going from N to S, so is in the positive x-direction at the wire. Remembering the current is downwards ($-z$) into the diagram, the force is evidently in the direction *perpendicular both to the wire and the magnetic field*.

Definition of Magnetic Field

The magnetic field strength $\vec{B}$ is defined by this force: for a uniform field, straight wire increment $\overrightarrow{d\ell}$,

$$\vec{F} = I\overrightarrow{d\ell} \times \vec{B}.$$

This result is well-established experimentally for any angle between the wire and the field, and in particular for a current running parallel to the field there is zero force.

This equation fixes the unit of magnetic field: for $\vec{F}$ in Newtons, I in amps, $\vec{B}$ is in *Teslas*.

Force on Any Current in a Constant Field

It is found experimentally that the total magnetic force on any wire carrying current I in a constant magnetic field $\vec{B}$ is the sum of terms $d\vec{F} = I\overrightarrow{d\ell} \times \vec{B}$.

For a constant magnetic field, for any shape wire going from $\vec{r}_1$ to $\vec{r}_2$,

$$\vec{F} = I\left(\int_{\vec{r}_1}^{\vec{r}_2} \overrightarrow{d\ell}\right) \times \vec{B} = I\left(\vec{r}_2 - \vec{r}_1\right) \times \vec{B},$$

The little $\overrightarrow{d\ell}$ vectors are head to tail, they all add to give the straight line from $\vec{r}_1$ to $\vec{r}_2$.

This means that for a closed loop of current in a constant field there is no net force—but there *is* in general a couple acting, as we'll soon discuss.

15 Magnetism II

Force on an Electric Charge Moving in a Magnetic Field

We've already discussed the experimentally well-established force on a current element in a magnetic field, recall that for an increment $\overrightarrow{d\ell}$ of current-carrying wire it was $\vec{F} = I\overrightarrow{d\ell} \times \vec{B}$.

If the linear charge density in the wire is λ coulombs/meter, and the charge is moving at $\vec{v}$ along the wire, then the force $\vec{F}$ on the charge q in the increment $\overrightarrow{d\ell}$ of wire, $q = \lambda d\ell$, is (using $I = \lambda v$)

$$\vec{F} = I\overrightarrow{d\ell} \times \vec{B} = \lambda d\ell \vec{v} \times \vec{B} = q\vec{v} \times \vec{B}.$$

(Of course, $\overrightarrow{d\ell}$ and $\vec{v}$ are parallel vectors.)

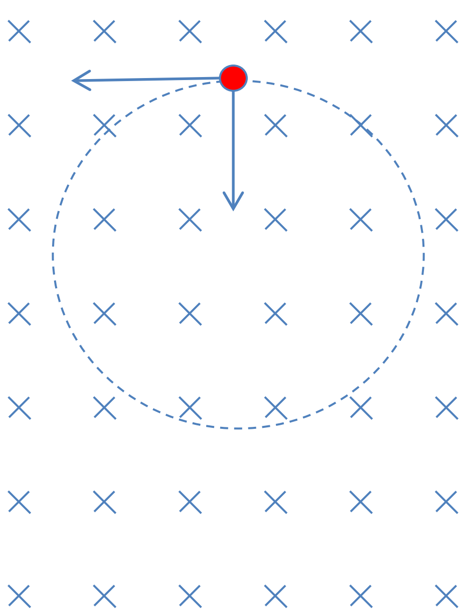
This, then, is the force on a charge q moving at velocity $\vec{v}$ in a magnetic field $\vec{B}$.

The force is *perpendicular to the direction of motion* at all times, so can do no work:

$$\vec{F} \cdot d\vec{s} = \vec{F} \cdot \vec{v}dt = q\left(\vec{v} \times \vec{B}\right) \cdot \vec{v}dt = 0.$$

Exercise: Suppose you have two parallel long wires carrying identical currents. They will attract each other, and accelerate towards each other. If the magnetic force can't do any work, how does this happen?

Hint: take the simplest possible picture—think of electrons going down a super pure metal wire, so the only constraint is that they stay in the wire by bouncing off the sides as they move down. How will switching on a magnetic field alter this picture?



Motion in a Uniform Magnetic Field

First, if the particle is moving parallel to the magnetic field it will feel no force and so continue at constant velocity.

Second, if it initially moving perpendicular to the uniform magnetic field, it will feel a sideways force proportional to its speed and will move in a circle, say radius $\vec{r}$, force $q\vec{v} \times \vec{B}$ pointing towards the center:

$$qvB = \frac{mv^2}{r}.$$

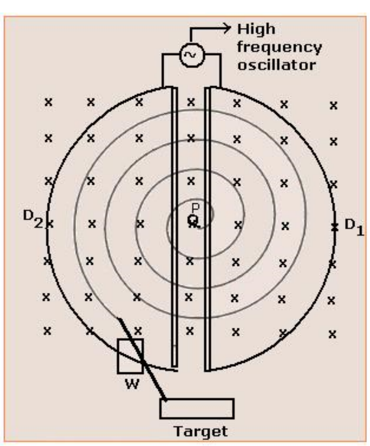
It follows immediately that the time for one circle is

$$T = 2\pi r / v = 2\pi m / qB$$

independent of the size of the circle!

This independence makes the cyclotron accelerator possible.

Proton in a Cyclotron



The two "D"s are hollow D-shaped metal boxes, open along the straight part.

The circling protons go back and forth.

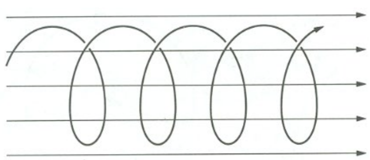
The oscillator alternates the relative voltages of the D's, so as a proton goes from one to the other it is attracted and accelerates, going into a larger, faster circle—*but with the same period*—each time.

If the proton reaches *relativistic* speeds, its mass increases and the circling time changes, recall $T = 2\pi r / v = 2\pi m / qB$.

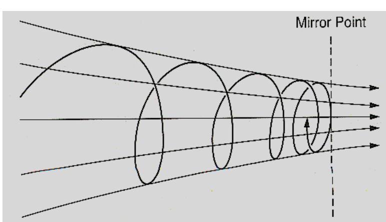
Still, in 1939 a 60-inch cyclotron at Berkeley accelerated deuterons to 16 Mev, and this was used in secret in the Manhattan Project to bombard Uranium and produce the Plutonium used in the "Fat Man" bomb, not declassified until 1948.

The relativistic mass circling time can be handled by having an oscillator with gradually decreasing frequency to match the mass increase. This is a synchrocyclotron: the problem is that now the particles must move in a tight group, whereas in the cyclotron particles could be fed in continuously.

Charged Particle in a Magnetic Field

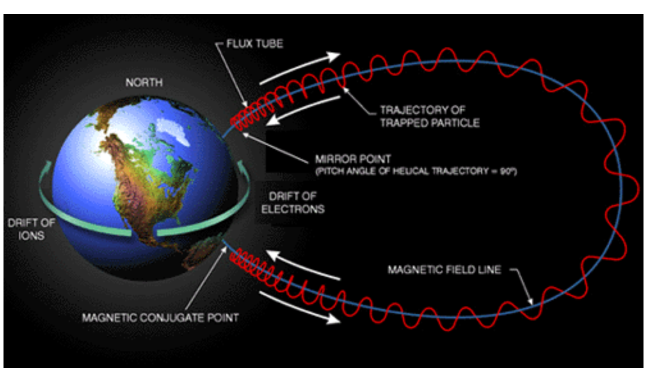


If the initial velocity is not perpendicular to the field, the motion in constant field will be circular plus a constant velocity parallel to the field—a helix.

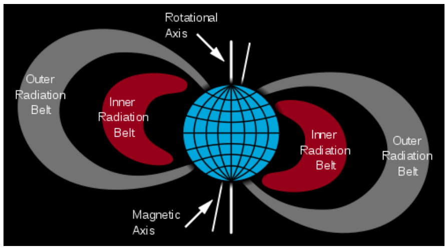


If the field is becoming stronger in the direction of motion, the helix gets tighter, and finally reverses. This is a *magnetic mirror*, used to confine plasmas in prototype fusion reactors. The slope of the field lines gives a "backward" component to the magnetic force.

Large-Scale Magnetic Confinement



The van Allen radiation belts are filled with charged particles moving between two magnetic mirrors created by the Earth's magnetic field. The outer belt is mostly electrons, the inner one mostly protons.



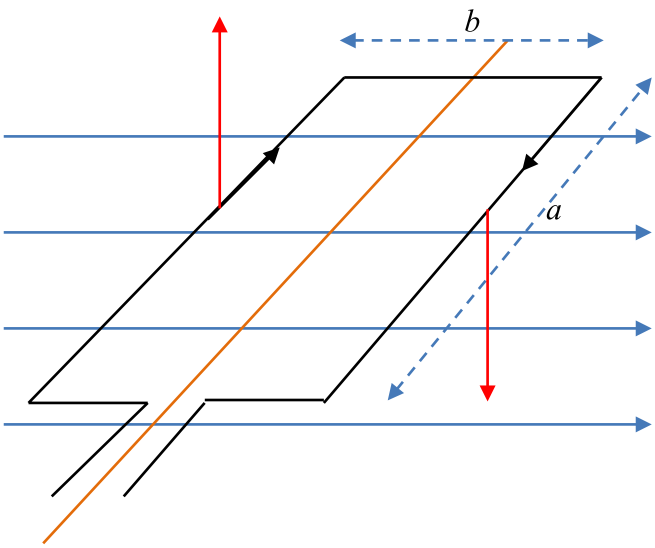
Physics 2415 Lecture 16: Magnetism III

Michael Fowler UVa

Torque on a Current Loop

This is the driving force for most electric motors, and, acting in reverse, the current generator for dynamos. It is also the basis for almost all pre-digital measuring devices: voltmeters, ammeters, etc.

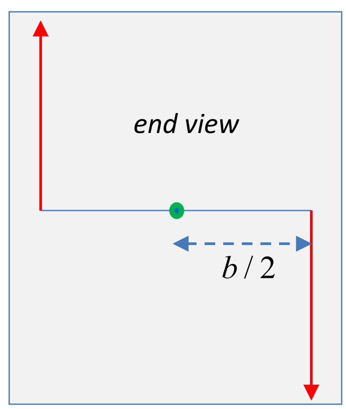
We begin with an $a\times b$ rectangular loop, horizontal, in a uniform magnetic field with field lines parallel to the end sides of the loop.

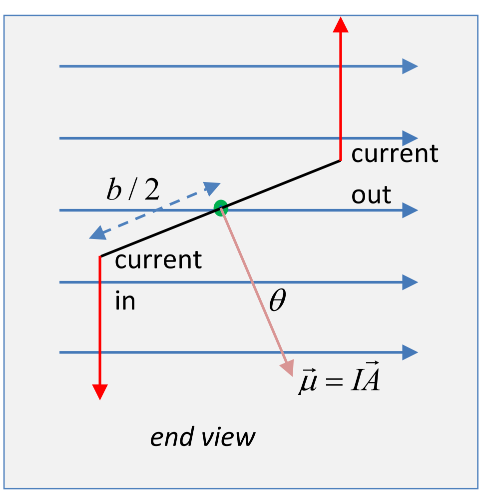


The forces on the other sides are vertical as shown, with magnitude $\left|I\vec{\ell}\times\vec{B}\right| = IaB$, and torque about the axis:

$$\tau = IaBb/2 + IaBb/2 = IabB = IAB$$

where $A = ab$ is the area of the loop.





Exercise: This formula (in terms of the loop area) works for any flat loop, not just rectangular. Try proving it!

Current Loop at an Angle

Note: for a coil with *N* turns, just multiply the single-loop result by *N*.

The current loop has a magnetic field resembling that of a short bar magnet, we define the *direction* of the loop area vector $\vec{A}$ (perpendicular to the wire loop) as that of the semi equivalent bar magnet, the magnitude of the vector $\vec{A}$ being the area.

Generalizing the result of the previous section to the case where the area vector $\vec{A}$ is no longer parallel to the magnetic field, the torque becomes (see figure)

$$|\vec{\tau}| = IAB\sin\theta,$$

meaning the loop has a dipole moment

$$\vec{\mu} = I\vec{A}$$

and as usual

$$\vec{\tau} = \vec{\mu} \times \vec{B}.$$

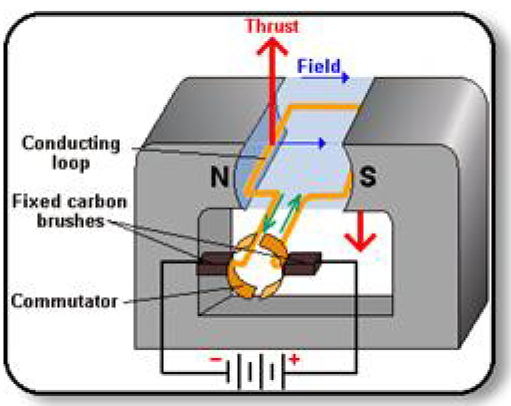
Note that the formula $\vec{\mu} = I\vec{A}$ is good for any flat loop.

Current Loop Potential Energy as Function of Angle to Field

The work done in turning the loop through incremental angle $d\theta$ is $\tau d\theta$, so, taking the zero of potential energy to be at $\theta = \pi / 2$, the potential energy at arbitrary θ is the work needed to get there,

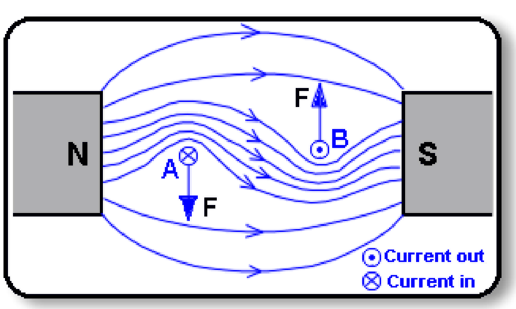
$$U = \int \tau d\theta = \int IAB\sin\theta d\theta = -\mu B\cos\theta = -\vec{\mu} \cdot \vec{B}.$$

Basic Electric Motor: the Commutator



It's just the loop in a magnetic field again, but with one crucial addition: the *commutator*.

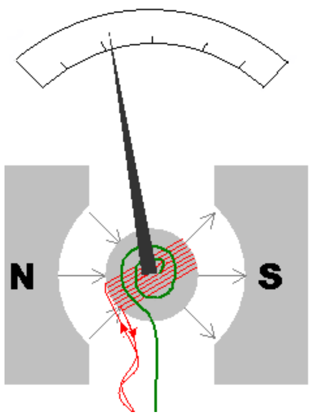
As the loop rotates (envision it as a short bar magnet attracted by the poles of the big magnet) the commutator *switches the current direction* (notice it's made of two half-circles) and therefore switches the loop's poles, so that the loop always feels a torque in the same direction (or zero), and continues to rotate.



As we've discussed earlier, Faraday pictured the magnetic field lines as elastic, naturally trying to shorten themselves (and also repelling each other sideways): this helps explain the force.

Exercise: Sketch the magnetic fields from (1) the permanent magnet and (2) the current in the wire independently, then see how adding them gives a picture like this.

To see a really simple motor, click here.



Galvanometer

The galvanometer measures the torque on a small coil in a magnetic field by balancing it against a curly spring (see figure). The coil is wound around an iron core to concentrate the field, and also to keep the coil in the same strength field when it turns within the angular limits of the instrument.

Trivia: Ampère named the galvanometer in honor of Galvani, the first person to detect a current, using frogs' legs (they twitched), years before the magnetic field from a current was detected with a compass.

Predigital voltmeters and ammeters are essentially all galvanometers. In the ammeter the current to be measured goes directly through the instrument. In contrast, the voltmeter is wired between two points to detect their potential difference by letting a very small current through the meter, so as not to impact the system significantly.

Thomson's Measurement of *e*/*m* for an Electron

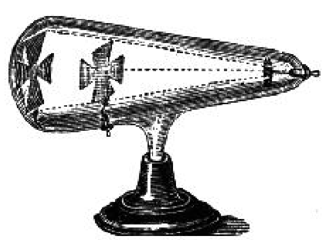


Fig. 515 Linear propagation of cathode rays; shadow formation

We discussed Thomson's experiment in lecture 11: electrons emitted by a heated wire cathode are accelerated by a high voltage in a vacuum tube and strike a phosphor-coated screen, leaving a shadow of any object (the anode here) in the way.

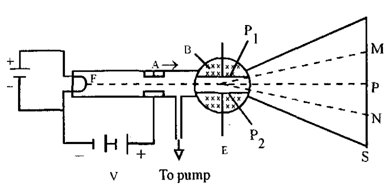


Fig. 9.2 J.J. Thomson's Experiment

Thomson narrowed the cathode rays to a pencil, which then passed between parallel charged plates (like a capacitor) P_1, P_2 (see figure) creating an area of uniform vertical electric field E. At the same time, current-carrying vertical coils were placed on each side of the tube to provide a uniform horizontal *magnetic* field B perpendicular to the ray direction, in the same region.

On entering the space between the plates, moving at speed v, the electron will be subject to a total vertical force, electric + magnetic, of $eE + evB$. Adjusting plate voltage or coil current until the electron goes through *without deviation*, its speed will be given by

$$v = E / B.$$

We know the original accelerating voltage V, and $eV = \frac{1}{2}mv^2$, so, having found the velocity v, we can now find e/m.

Millikan's Oil Drop Experiment

To measure just the charge of an electron, it doesn't work to balance the electric force with a magnetic force, both depend on the charge. It is necessary to balance the force eE with a *nonelectrical* force. The obvious candidate is gravity, and for the force on a single electron, we need the gravitational force on a small object. Millikan (at Chicago, 1908) chose the tiny oil drops emitted by a mist spray bottle, used for example for perfume. The cloud of mist generated takes some time to settle under gravity, because the weight of a (spherical) small drop is essentially balanced by the viscous friction air resistance as it falls. For a drop of radius r, the air resistance at speed v is $6\pi r\eta v$ where η is the air's viscosity (known), so measuring v and using $6\pi r\eta v = \frac{4}{3}\pi r^3 \rho g$ (ρ the oil density) the radius can be found, and hence the weight.

The procedure, then, is to generate a small cloud of spray, then use a microscope to find a drop falling at an easily measurable rate, and thus calculate its radius and hence its weight.

(*Aside: this was evidently more accurate than measuring the size of the drop by observation. Discuss*.)

Next, a vertical electric field is turned on and adjusted until the oil drop stops falling, becoming stationary. At this point, the weight is balanced by the electrical force qE so q can be found. The experiment is repeated many times, and it is found that the measured charge is always a whole number times a basic unit: $q = ne$, n an integer and $e = 1.6\times 10^{-19}$ coulombs. This then is the electron charge.

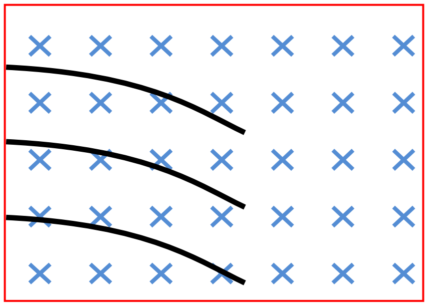
More recent update: After the invention of quark theories in the seventies, millions of dollars were spent repeating Millikan's work to search for free quarks (meaning not inside another particle), which would have charges one-third or two-thirds the electron charge. No free quark was ever found.

Exercise: Taking the oil density to be approximately that of water, what is the radius of a droplet in balance with one excess electron in a field of 1000 volts/meter? Look up air's viscosity to find its approximate rate of descent if the electric field is switched off, using the (Stokes') viscosity drag formula given above.

Hall Effect

The Hall effect was discovered by a Johns Hopkins graduate student, Edwin Hall, in 1879. He asked himself whether the force felt by a current-carrying wire in a magnetic field was a force on the wire or on the current. Remember this was before the discovery of the electron, so the concept of the electrical current was pretty vague.

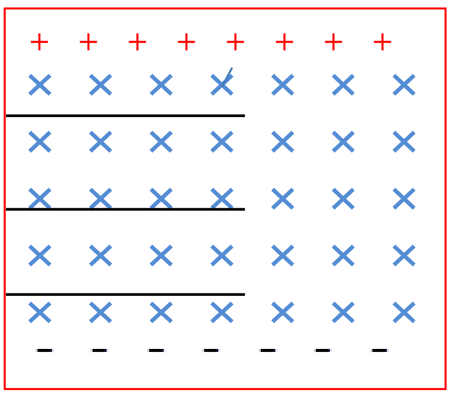
The essence of the Hall Effect can be understood with a simple model. Suppose with zero magnetic field charged particles are fed into the left-hand side of a conductor and a horizontal electric field keeps them moving to the right. (We're looking at an *average* of many particles, a single particle will follow a complex path with many collisions, see Drude.)



Next we switch on a magnetic field perpendicular to the screen, pointing inwards. The particles will experience a force $q\vec{v}\times\vec{B}$ and for negatively-charged electrons this will deviate them downwards. But this pattern won't last long: negative charge will pile up along the bottom edge, generating a repulsive electric field which will eventually exactly compensate the magnetic force, this is called the Hall field, and written

$$E_H = v_\text{d}B$$

where v_d is the average, or drift, speed of the charged particles.



The Hall voltage, or emf, is the potential difference between the top and bottom of the strip. For width w, this is

$$\mathcal{E}_H = E_H w = v_\text{d}Bw.$$

Exercise: It was not known at the time of this experiment if the current was negative particles moving to the right or positive particles moving to the left. But the experiment could distinguish between these models. Explain why.

Mass Spectrometer: Velocity Selection and Identification

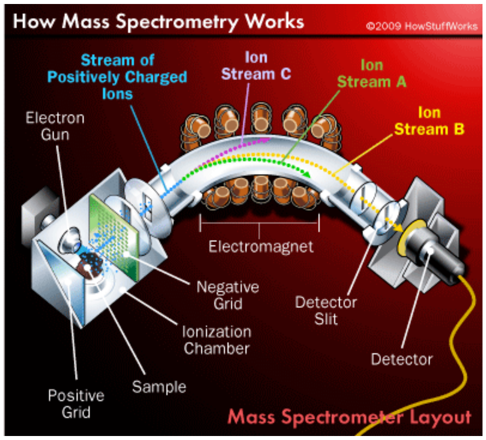
In the above discussion of Thomson's experiment to measure *e*/*m*, for a stream of particles moving in the x-direction at speed v, if there is an electric field of strength E in the y-direction and a magnetic field of strength B in the z-direction such that $\vec{E}+\vec{v}\times\vec{B}=0$ there will be no net force and the stream of particles will not be deviated. Note that this condition does not depend on the mass or the charge of the particles.

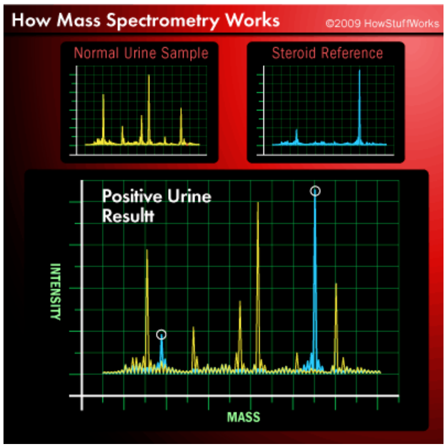
This means that if we send a stream of different kinds of particles, different masses, charges, velocities, down a narrow tube with these sideways electric and magnetic fields, only those with speed E/B will get through, the others will deviate and hit the sides.

Once we have a stream of particles all at the same velocity, we direct them into a perpendicular magnetic field (now no electric field) and they will circle with radial acceleration $v^2/r = qvB/m.$ If we detect them after half a circle, the half circle path radius will be proportional to the particle mass, so we can read off the proportion of different particle masses in the stream: this is termed Mass Spectrometry.

One use of mass spectrometry is carbon dating. Most elements have several isotopes: the nuclei of course have the same number of protons, but different numbers of neutrons. The isotopic ratio can change with time if one of the isotopes is radioactive and so decays. This is the case with the isotope carbon-14, continually produced from nitrogen (and cosmic radiation) in the upper atmosphere, but when absorbed into living tissue it decays with a half-life around 5700 years, so its fraction is good for dating in the range 500 – 50,000 years.

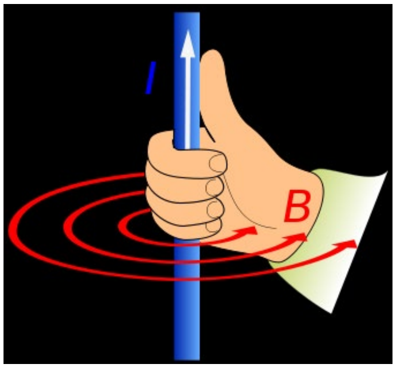
Mass spectrometry also works for molecules, an example being drug detection in a urine sample.





17 Sources of Magnetic Field I

Magnetic Field from a Current in a Long Straight Wire



From many experiments, the lines of magnetic force are circles around the wire, direction determined by the right-hand rule.

The field *strength* is proportional to the current, and inversely proportional to distance from the wire.

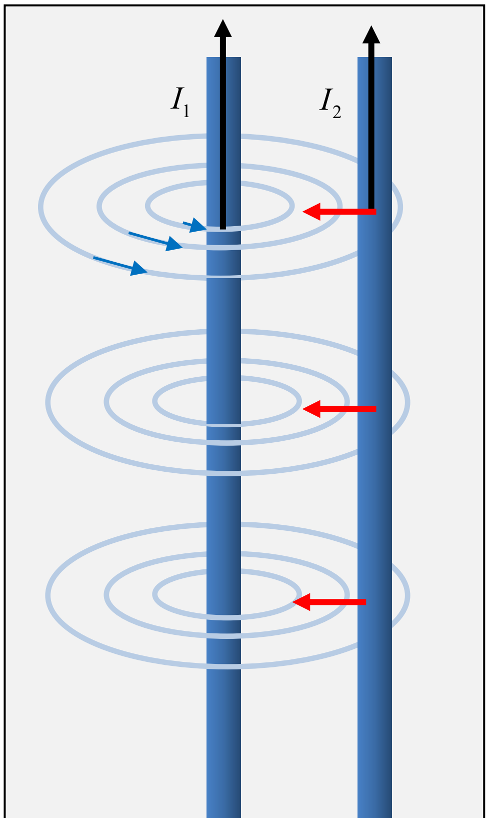
$$B = \frac{\mu_0}{2\pi}\frac{I}{r}, \quad \mu_0 \cong 4\pi \times 10^{-7} \text{ Tesla.m/A.}$$

μ_0 Update

In an earlier version of these notes, we wrote $\mu_0 = 4\pi \times 10^{-7}$ exactly, and that was true at the time—but things have changed. From 2019 on, the speed of light, the electron charge and Planck's constant have been given specific numerical values (effectively, this defines the units of length, etc., the unit of time having been defined in terms of certain frequencies of the Caesium atom) and the result is that μ_0

changed from its previously defined value by about one part in ten billion. So for this course, and likely the rest of your life, you can stick with the old value—but be aware of this trivium.

Force Between Parallel Wires



Showing field from I_1, force on I_2.

The field from current I_1 is $B=\frac{\mu_0}{2\pi}\frac{I_1}{r}$, circling the wire, and the current I_2 will feel a force $I_2\vec{\ell}\times\vec{B}$ per length ℓ, so the force per meter on wire 2 is

$$F=\frac{\mu_0}{2\pi}\frac{I_1 I_2}{d}$$

towards wire 1 and wire 1 will feel the opposite force.

Bottom line: ***Like currents attract.***

Definitions of the Ampere and the Coulomb

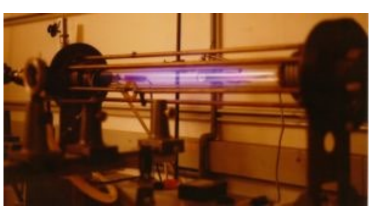
The traditional definition of the ampere, the unit of current, is based on the two-wire scenario above, two long equal parallel *one amp* currents one meter apart feel an attractive force 2×10^{-7} N/m.

The unit of charge, the *Coulomb* is the charge flow per second in a one amp current.

As mentioned above, the units have been redefined, but the change is around one part in ten billion so will not concern us.

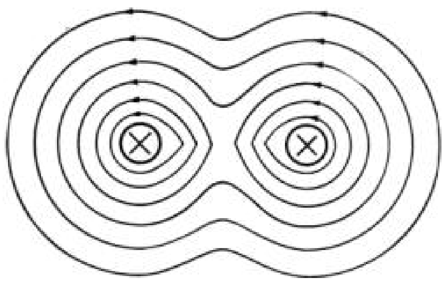
Like Currents Attracting

This is a piece of copper pipe: lightning ser hrough it, the parallel currents attracted each other and pulled the pipe w a very hot central volume.



The same thing happens on sending a large current through a plasma, and intense heat is generated. This is one possible scenario for raising the temperature of small nuclei sufficiently to trigger fusion. Unfortunately, plasmas have many instabilities under these conditions, and it seems decades will still be needed to make this a practical power source.

Magnetic Field Lines for Parallel Wires

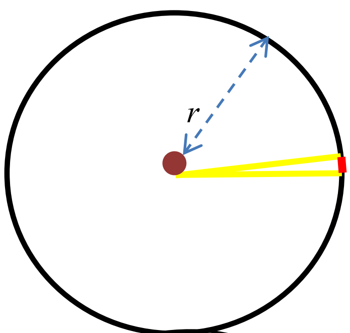


The magnetic field at a point is the vector sum of the two fields circling the wires.

Exercise: check the diagram on the left (for equal currents) by first sketching the two sets of circles (one for each wire) then use a different color for the vector sum in a few places, to see how this pattern emerges. What's going on in the middle?

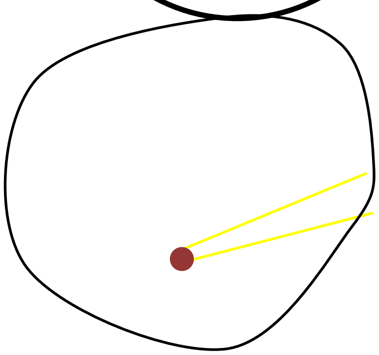
Finally, sketch the field lines if the currents are *opposite*. Make clear what happens in the middle.

Introducing Ampère's Law



Consider a current I in a long wire perpendicular to the screen, and the integral $\oint \vec{B} \cdot d\vec{\ell}$ around a circle of radius r as shown. Taking $B = \mu_0 / 2\pi r$, and $d\ell = rd\theta$ (the small red vector), the integral is just over θ from zero to 2π and

$$\oint \vec{B} \cdot d\vec{\ell} = \frac{\mu_0}{2\pi r} \oint r d\theta = \mu_0 I.$$



Suppose now we take a *noncircular* contour for the integral. The increment $\vec{B} \cdot d\vec{\ell}$ only has a contribution from the component of $d\vec{\ell}$ which is parallel to $\vec{B}$, and has length $rd\theta$, so we get the same result for the integral.

This is even still true if we draw the curve in three-dimensional space, because the added dimension (perpendicular to the screen) is perpendicular to $\vec{B}$ so makes no contribution to $\vec{B} \cdot d\vec{\ell}$.

Exercise: **Important!** Do the same exercise but with the wire *outside* the curve.

Prove the answer is *zero*. *Hint*: Track what happens to θ as you go around once, check $\oint d\theta$.

Ampère's Law

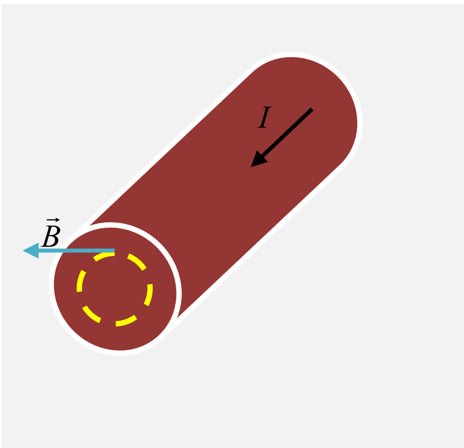
From the two cases discussed above, we can see that for a magnetic field from many long straight wires in arbitrary directions,

$$\oint \vec{B} \cdot d\vec{\ell} = \mu_0 I_{\text{encl}},$$

where I_{encl} counts only currents that penetrate a surface roofing the integration curve.

This is *Ampère's law*, and in fact is true for any collection of time-independent currents, well-verified experimentally. (Our "proof" above is only for a collection of straight-line currents. A general proof is not difficult, but needs a bit more calculus.)

Field Inside a Wire



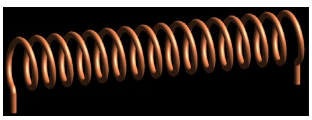
Apply Ampère's law to the dashed circular path of radius r inside the wire as shown. From symmetry (and no monopoles), the field must be tangential.

The surface "roofing" this path has area πr^2, the whole wire has cross-section area πR^2 so the current flowing through the path is Ir^2 / R^2, and Ampère's law gives

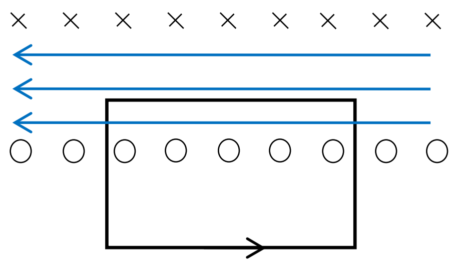
$$\oint \vec{B} \cdot \overrightarrow{d\ell} = 2\pi r B = \mu_0 I r^2 / R^2,$$

$$B = \frac{\mu_0 I}{2\pi} \frac{r}{R^2}.$$

Field Inside a Solenoid



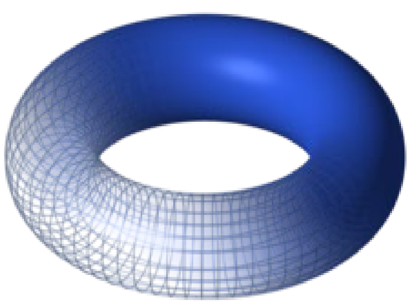
Take a rectangular Ampèrian path as shown below. Assume the external magnetic field negligible, and the field inside parallel to the axis (a good approximation for a long solenoid). For current I, n turns/meter,



$$\oint \vec{B} \cdot \overrightarrow{d\ell} = B\ell = \mu_0 n I \ell,$$

$$B = \mu_0 n I.$$

Magnetic Field of a Toroid



A toroid here is equivalent to a solenoid with the axis turned into a circle so the two ends connect. This is a promising design for containing a hot plasma, unlike the "magnetic bottle" where charged particles can escape at the ends.

Notice the current-carrying wires spiraling around the surface.

To find the field, imagine slicing the donut to get maximal flat surface area, the wires intersect this surface in two concentric circles, say currents coming up on the inner circle, down on the outer circle.

From symmetry, the lines of magnetic field must themselves be circles centered on the main axis, and the field must have the same strength all the way round.

For the integral around a circle of radius r,

$$\oint \vec{B} \cdot \overrightarrow{d\ell} = 2\pi r B = \mu_0 N I$$

where N is the number of times the wire penetrates the circular disk having the circular contour as its boundary, meaning the number of times the wire circles the contour.

If the circular contour is inside the toroid, it contains the inner circle, so $B = \dfrac{\mu_0 N I}{2\pi r}$, notice this is not uniform, unlike the linear solenoid. If the circular contour is outside the solenoid, both up currents and down currents penetrate the circular area, cancelling, and there is no field. In fact, it is easy to show there is no field except within the toroid volume.

18 Sources of Magnetic Field II

Magnetostatics: the Biot-Savart Law

Finding the magnetic field from a steady current distribution is called *magnetostatics*, in analogy with electrostatics, which is finding the electric field from a stationary charge distribution. But there we had a very definite prescription: we knew the inverse-square field from a point charge, and we used the principle of superposition, adding together the fields from all the charges, to find the total field.

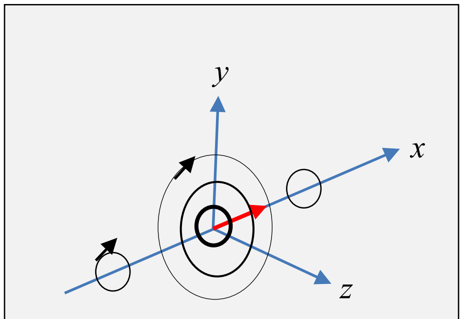
When Ampère was doing his experiments, in Paris in the 1820's, his colleagues included some of the world's best mathematicians, and they set about doing for magnetostatics what had already been accomplished in electrostatics: they looked for a formula for the magnetic field from a little bit of current, so that using superposition they'd be able to find the field from any distribution of currents by adding all the elements, just like the electric field from many point charges.

In fact, two of them succeeded in finding a formula that worked, but it's a very strange formula. It's called the *Biot-Savart* law, and here it is:

The magnetic field at $\vec{r}$ from an infinitesimal length $d\vec{\ell}$ of wire carrying current I at the origin is

$$d\vec{B} = \frac{\mu_0}{4\pi} \frac{I d\vec{\ell} \times \hat{r}}{r^2}$$

Notice it is inverse-square, like electrostatics (remember $\hat{r}$ is a unit vector).



Field from current element $I\vec{d\ell}$

It's worth thinking about this field a little. Take $d\vec{\ell}$ at the origin and pointing in the x-direction. What are the field lines in the plane $x=0$? (This plane includes the y and z axes.) They are circles, curling around as given by the right-hand rule. But they're not like the field from a wire along the x-axis: the field strength from this little current element goes down as the inverse square, evidently Ampere's law doesn't work for this current. Actually on the x-axis, anywhere, the field is zero, and everywhere else it's perpendicular to the x-axis and circling around it.

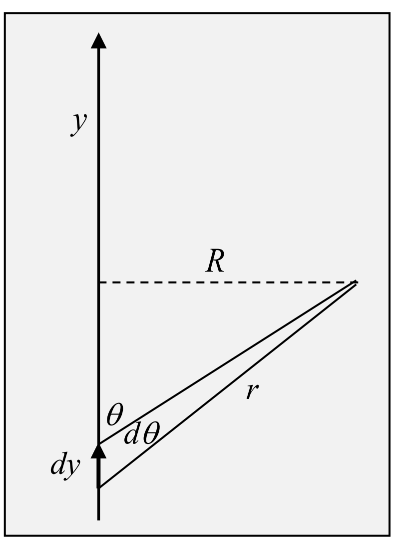
In one way at least, this formula is just a mathematical trick: you can't physically have a little element of steady current, opposite sign charges would be piling up at the two ends. Such an element only has meaning as part of a complete circuit.

The forces between two current elements *aren't even equal and opposite***.** Consider this by taking a second current element, at $\begin{pmatrix}1 & 0 & 0\end{pmatrix}$ pointing in the y-direction. It will feel no force from the first current element, but the field from the second current element is certainly nonzero at the origin.

(*Another point*: suppose these two "current elements" are just nonrelativistic moving charged particles. If we assume the Biot-Savart law is still good, apparently Newton's Third Law doesn't work for the magnetic interaction? The answer is that the electric and magnetic *fields* carry energy and momentum—not just the particles—so *total* momentum can be conserved even if the forces between *particles* are not equal and opposite. But this is too complicated to analyze here.)

So this is a strange formula, but it works. Consider a straight finite stretch of wire, part of some circuit. What's the field at distance R?

The Biot-Savart rule tells us



$$dB=\frac{\mu_0}{4\pi}\frac{Idy\sin\theta}{r^2}$$

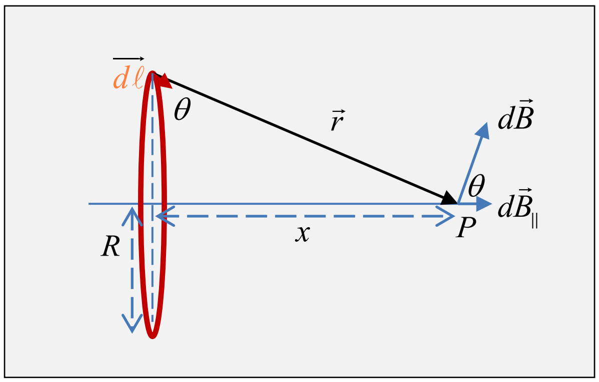
and from the discussion above all the dB's point into the screen, so we just integrate over the (finite) length of wire we're considering. It's easiest to switch variables from y to θ, notice $\cot\theta = y/R$, so $dy = R\operatorname{cosec}^2\theta d\theta = r^2 d\theta/R$, the integral just becomes

$$B=\frac{\mu_0 I}{4\pi}\int_{\theta_1}^{\theta_2}\sin\theta d\theta/R=\frac{\mu_0 I}{4\pi}\left(\cos\theta_1-\cos\theta_2\right)/R.$$

This is done in the book for an infinite wire—but it works for a finite length too, provided it's part of a circuit, so charge isn't piling up. For example, you could find the field from a square coil.

For a round coil, though, the integral isn't that easy, *except* for the important case of the field on the axis, let's work it out.

We show the ring of current as copper-colored, and focus on an increment $I\overrightarrow{d\ell}$ at the top of the loop (this little vector is perpendicular to the screen/page). At a point P on the ring's axis distance $\vec{r}$ from the current increment, the magnetic field contribution from this increment is perpendicular to $\vec{r}$, and has a component parallel to the x-axis

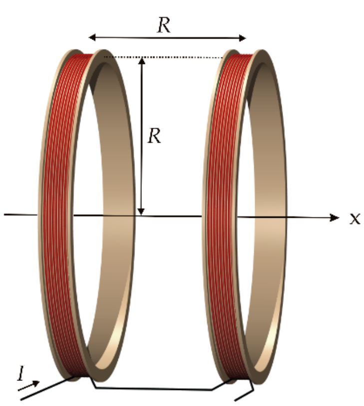


$$dB_{\parallel} = \frac{\mu_0 I d\ell}{4\pi r^2}\cos\theta$$

Adding the contributions from all around the circle, the components perpendicular to the x-axis cancel out by symmetry, those along the x-axis add to give

$$B = B_{\parallel} = \frac{\mu_0 I 2\pi R}{4\pi r^2}\frac{R}{r} = \frac{\mu_0 I}{2}\frac{R^2}{\left(R^2+x^2\right)^{3/2}}.$$

At distances $x \gg R,\ B \cong \frac{\mu_0}{2\pi}\frac{\pi R^2 I}{x^3}$, which is the field from a dipole of strength $\pi R^2 I = AI$, with A the area of the loop, a formula that turns out to be good in this limit for any shaped loop.



Helmholtz Coils

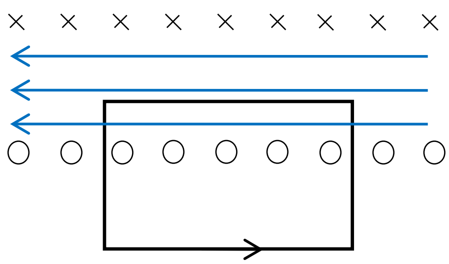
Two identical circular coils of radius R are distance R apart along their common axis, as shown. This provides an experimentally useful very uniform field at the midpoint of the setup, the field only increases 7% on going from that center point to the plane of a coil. This configuration is sometimes used to cancel the Earth's field.

If the currents in the two coils are opposite, the magnetic field at the center point is zero and linear in x to a very good approximation, this is useful in magneto-optical traps.

Exercise: check these field facts.

Magnetic Field inside a Long Solenoid

The dots and crosses are the loops of current going in and out of the paper. The blue represents the fairly uniform magnetic field inside the solenoid, compared with which the field outside is negligible. We take the current I, and the number of coil turns per unit length n.



Then from Ampère's Law integrated around the black rectangular contour, of length L,

$$\oint \vec{B}\cdot\overrightarrow{d\ell} = BL = \mu_0 nIL,$$

So $B = \mu_0 nI$

and is uniform across the cross section, provided we can neglect end effects (so a very long solenoid).

Diamagnetism and Paramagnetism: Permeability

Diamagnetism is a molecular version of Lenz' law. Orbiting electrons in atoms and molecule are little currents, and when the magnetic field going through a current loop changes, the current itself changes in a way to minimize the total change in magnetic field. In other words, if the magnetic field through a diamagnetic solid changes, the solid generates its own field to lessen the change.

A quantitative measure of diamagnetic response is the permeability, denoted by μ. This is a measure of the material's response to a magnetic field, so if the long solenoid discussed just above is filled with material the field inside will be $B = \mu nI$.

The constant μ_0 is often called the *permeability of the vacuum*.

For many substances μ is very close to μ_0, and a convenient parameter is the magnetic susceptibility

$$\chi_m = \frac{\mu}{\mu_0} - 1.$$

For diamagnets, the effect is usually small, $\chi \sim -10^{-5}$ except bismuth, $\chi \approx -1.66\times 10^{-4}$, and one big exception: superconductors, which exhibit the *Meissner effect*: on putting one into a field, surface currents appear and the field (below a certain strength) cannot penetrate the superconductor, so $\chi = -1$.

All solids have some diamagnetic response, but if some of the atoms also have nonzero magnetic moments, there is also a (usually stronger) *paramagnetic* response, as the molecular magnetic moment tends to align with the applied field. For paramagnetic materials $\chi \sim 10^{-5\pm 1}$.

For paramagnets, the response to an external field can be found by analyzing a single spin. The analysis breaks down at very low temperatures when enough atoms are aligned with the field to contribute a sufficient extra overall magnetic field. This was at first thought to explain ferromagnetism, but, on running the numbers, in iron crystals this magnetic moment lining-up doesn't occur above a few degrees absolute, evidently something else is lining up the moments in iron. Read on.

Ferromagnetism: Domains

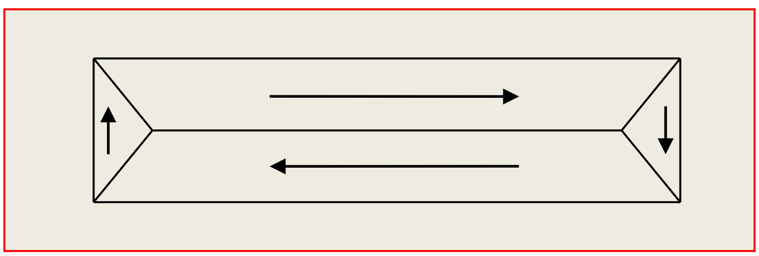
In a f*erromagnet*, the individual atoms are little magnets, but in contrast to a paramagnet, there are powerful *quantum mechanical* forces causing nearest neighbors to align magnetically. (So this alignment does *not* come from the much weaker magnetic dipole-dipole interaction.)

In particular, the atoms of Fe, Co and Ni (and rare earths) are little magnets: in the incompletely filled shell of electrons, the electron spins line up—and electrons are themselves magnets. (If you're familiar with quantum mechanics, the spins line up to give a symmetrical spin wave function, which means the spatial wave function must be antisymmetric, and that keeps the electrons from getting too close, so minimizing the repulsive electrostatic energy.)

So why isn't every piece of iron magnetic?

What actually happens is best illustrated by considering a very pure small crystal of iron called a whisker. On looking closely, it turns out that almost all atoms *are* aligned with their neighbors, but the crystal as a whole is divided into fully aligned regions, called *domains*, as shown below.

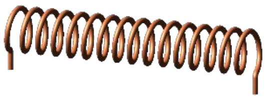
The reason is that this is the state of lowest energy. To create a magnetic field costs energy, and if all the atoms were aligned, there would be a strong dipole field in the surrounding space. The arrangement shown here generates very little external field. Of course, the domain walls cost some energy, so calculations are needed to find the optimum configuration.



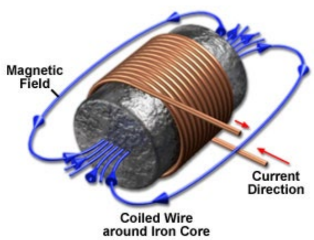
If this whisker is placed in an external magnetic field, the domain closest to parallel with the external field will grow at the expense of the others. This lowers the energy of the whisker in this new environment, just as a compass needle will swing around to align itself with an external field.

Actually this whisker is very pure iron, a single crystal, and so very soft, magnetically speaking. That means it readily responds to a change in external field, which in turn means that the boundaries between domains move easily. This is good for iron in the core of an electromagnet, but not desirable in a permanent magnet. There you need to be able to line up domains and make it hard for them to readjust. This will be the case if the single crystal is replaced by many small crystals having different axes, and also certain non-iron atoms, such as in an alloy, can pin the domain walls and make movement difficult. Obviously, to make a permanent magnet takes a more intense external field than is necessary for magnetizing soft iron temporarily.

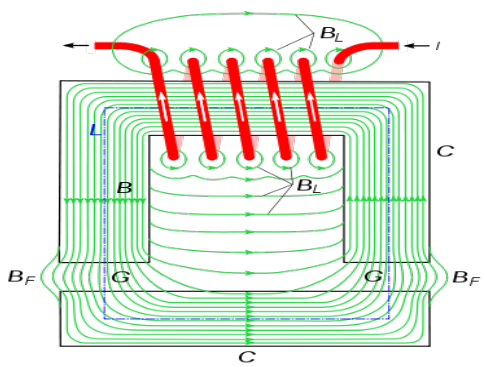
Using Soft Iron to Make a Strong Electromagnet



We've already seen that a way to concentrate the magnetic field from a current in a wire is to form a solenoid, which then resembles a bar magnet. A much stronger field can be achieved by filling the core of the solenoid with a soft ferromagnetic material. "Soft" in this context means a material very responsive to the field, the domain walls move readily to ensure a full magnetic response, that is, the atomic magnets in the material fully line up with the prevailing field. A soft iron core can increase the magnetic field in a solenoid by a factor of thousands (the factor μ/μ_0). Mumetal goes to hundreds of thousands, and there are alloys at one million. Mumetal is used to shield a hard disk from the field of the motor.



(*Note*: in some books, you might see the notation $B = \mu H$ where B is the physical magnetic field and H is the field that would be produced by the currents in the wires if no magnetic material were present. This is perhaps useful for engineering design, but we won't be using it.)



Here is a basic design for the type of magnet used to pick up pieces of wrecked cars, etc. Notice how the soft iron concentrates the field lines. The Π shape is the body of the magnet.

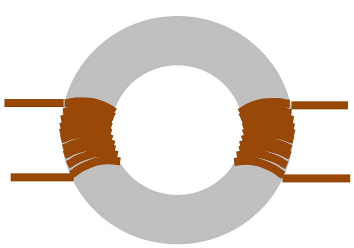
The bar along the bottom is the piece of car being picked up.

Click the picture for more details.

19: Magnetic Induction I

Faraday's Idea

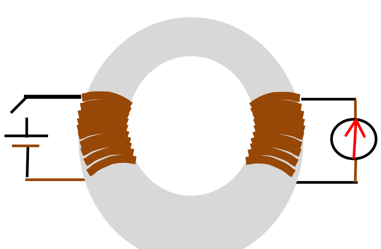
Faraday theorized that since an electric current could generate a powerful magnetic field, maybe a magnetic field could generate a current?



He tested this theory by winding *two* solenoids around the same doughnut shape of soft iron.

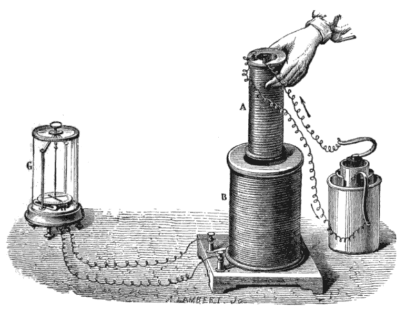
He ran a large current in one, looked for a current in the other—and didn't find it.

But he did find something!



He found a *transient* current appeared in the second coil *at the moment the current in the first coil was turned on*, then a transient *opposite* current when it was turned off.

Induced EMF



Faraday discovered that what he called "induced current" appeared in a coil whenever the external magnetic field through the coil was changing.

Here is one of Faraday's experiments as portrayed in an 1892 physics textbook "for advanced students". On the right is a battery, on the left a fancy galvanometer.

We say there is an induced *emf* driving this current, emf being short for "electromotive force", the "force" driving the current, the voltage. Other sources of emf are electric fields, and the chemical forces inside a battery.

Faraday's Experimental Findings about emf

For a coil of N loops close together, he found the induced emf to be N times that for one loop (meaning the current will be the same if there's negligible external resistance in the circuit).

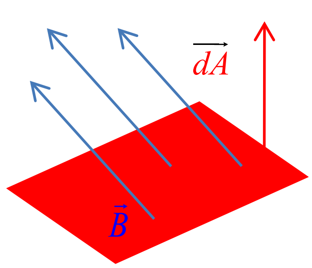
For a uniform magnetic field, the emf is proportional to the area of the loop.

It's proportional to the component of magnetic field perpendicular to the area.

It's proportional to the *rate of change* of field.

Faraday thought of the magnetic field lines as representing flow of some ethereal fluid, rather analogous to the electric field—but with one big difference. The electric field "fluid" flowed out of positive charges, into negative charges. In contrast, on looking at the magnetic field of a solenoid, for example, the north and south poles are *not* like positive and negative electric charge, they're illusions, the magnetic field lines do not end, they just circle around.

To quantify the interaction of the magnetic field with the loop, bearing this fluid picture in mind, a natural quantity to measure is what is the total magnetic flux (Latin for flow) through the loop, to be measured by putting a roof over the loop and measuring the total magnetic field through the area.

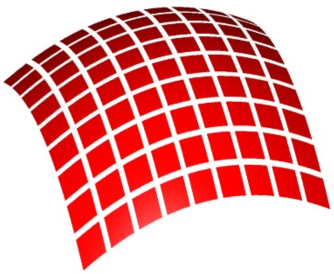


The flux through a small square with area $\overrightarrow{dA}$ is $\vec{B}\cdot\overrightarrow{dA}$.

The total magnetic flux through the surface bounded by the loop is written:

$$\Phi_B = \int \vec{B} \cdot \overrightarrow{dA},$$

summing the contributions from all the small squares.



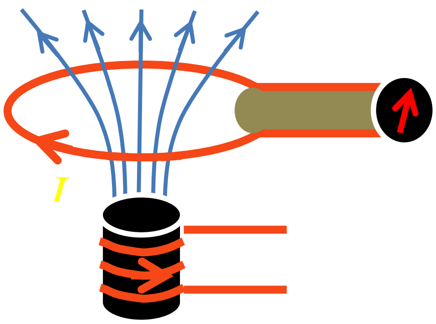
Recall now Faraday's first idea was to see if having a magnetic flux through a loop/coil generated a current, just as having a current generated a field.

It didn't—except when it changed.

Faraday's Induction Formula

He saw transient currents *when the flux through the loop changed*: the induced emf he measured as

$$\mathcal{E} = -\frac{d\Phi_B}{dt}.$$

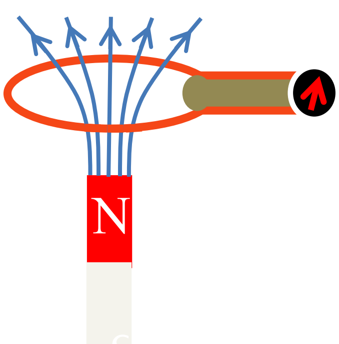


Lenz' Law

The sign of the induced emf is most simply found by applying *Lenz' law*, the induced current will generate a magnetic field opposing the change in field flux caused by the movement.

So if the current builds up from zero in the solenoid in the direction shown, a north pole is appearing at the top of the solenoid, so the current generated in the loop will have a north pole on its underside, to partially cancel the field from the solenoid.

Notice this also means there is a momentary *repulsive* force between the coil and the ring, if the ring is not attached, and the current is sufficient, the ring will jump. There are many YouTube demos, for example this.



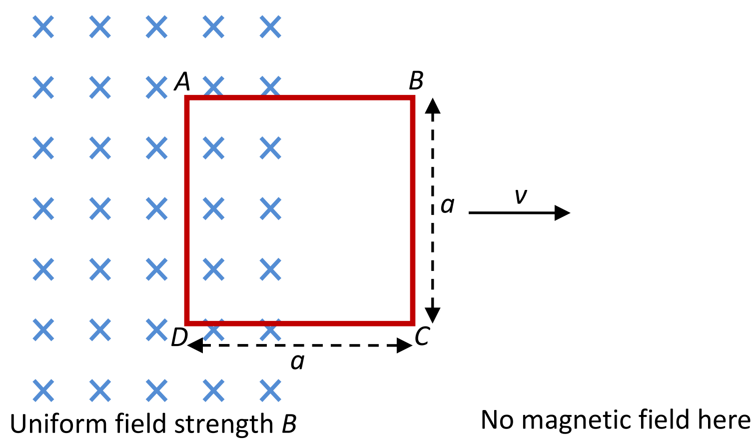
We could of course get the same effect by physically moving a magnet towards the coil instead of turning on the solenoid. It's useful to consider this alternative, because it makes explicit that as the response current builds up in the ring, it generates a repulsive force so it takes work to keep moving the magnet, this is the source of the energy needed to build up the current in the ring (which then dissipates as ohmic heat).

Exercise: Discuss the energy balance when turning on the solenoid.

20 Magnetic Induction II

Pulling a Square Loop out of a Magnetic Field

Let's assume for purposes of illustration that we have a magnetic field pointing inwards, and it ends suddenly at the plane $x = 0$ (which has to be an approximation, since Ampere's Law wouldn't be satisfied integrating $\vec{B} \cdot \overrightarrow{d\ell}$ around a loop perpendicular to this one—but it could be a good approximation. For example, we could be pulling the loop out of the strong field between the poles of a permanent magnet).



Pulling a square conducting loop out of a uniform magnetic field *B* at speed *v*

From Faraday's law, we know the emf $\mathcal{E}$ induced is

$\mathcal{E} = -d\Phi_B / dt$, so in this case $\mathcal{E} = avB$, and the current direction from Lenz' law is clockwise, to generate inward magnetic flux to replace some of that lost.

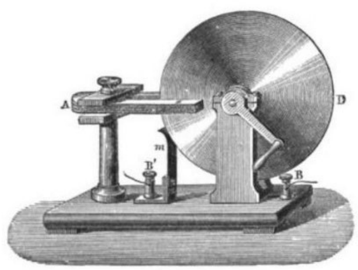
There's another way to understand this: consider the electrons in the leg *AD* of the loop. As a result of the loop being pulled sideways at speed *v*, they will feel a Lorentz force $\vec{F} = q\vec{v} \times \vec{B}$, driving them down the wire (since they're negatively charged). And since the leg *BC* is out of the field, there is no balancing force to prevent a current from flowing. Relative to the electrons in the wire, the *moving* magnetic field has generated an electric field along the wire of strength *vB*, corresponding to a potential difference, or emf, of *avB*.

*(*Footnote for anyone interested*: In fact I think this argument works in the general case. Visualizing the magnetic field a la Faraday, in terms of lines of force, suppose any loop of wire has the magnetic flux linking through it changing, because of the loop contorting, moving, or the field changing, or whatever. It seems clear from the Faraday picture that flux can only become unlinked by moving *across the wire* (which itself may be just a mathematical curve). If the relative motion of an infinitesimal part of the loop and the local magnetic field $\vec{B}$ is $\vec{v}$, there is an increment of potential from the Lorentz force $d\mathcal{E} = \vec{v} \times \vec{B} \cdot d\vec{\ell}$, and writing $\vec{v} = d\vec{x} / dt$ this is

$$d\mathcal{E} = \frac{d}{dt}\left(d\vec{x} \times \vec{B} \cdot d\vec{\ell}\right) = \frac{d}{dt}\left(d\vec{\ell} \times d\vec{x} \cdot \vec{B}\right) = \frac{d}{dt}\left(d\vec{A} \cdot \vec{B}\right)$$

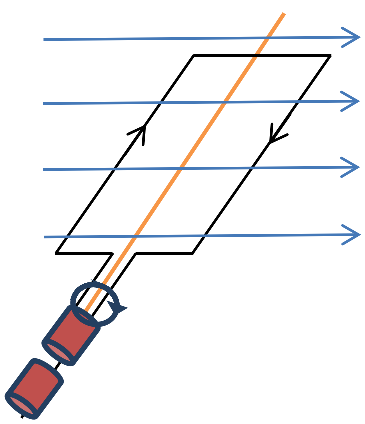
where $d\vec{A}$ is the element of area swept out by the wire's movement relative to the magnetic field. But $d\vec{A}\cdot\vec{B} = d\Phi$, the change in enclosed flux caused by this movement. One can write this $d\mathcal{E} = \vec{E}\cdot d\vec{\ell}$ where $\vec{E}$ is the electric field in the wire's frame of reference, and, adjusting signs suitable using Lenz' law, this gives $\int \vec{E}\cdot d\vec{\ell} = -d\Phi_B / dt$. What about the uniform magnetic field between close flat poles of a large electromagnet, when the current is increasing? I would visualize that as magnetic field flowing radially inwards, find the appropriate velocity, etc.)

Electric Generators



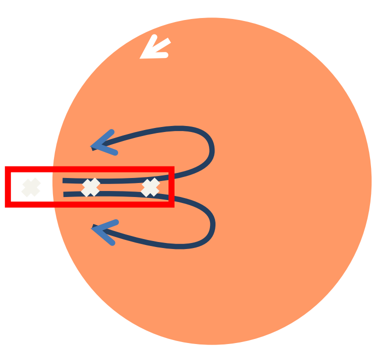
Here is a copy of the first electric generator, constructed by Michael Faraday in 1831. A is the magnet; B, B' the terminals. On rotating the copper disc, an emf is generated in the region between the poles of the horseshoe-like magnet A. Taking that field to be into the screen, and the disc spinning anticlockwise, an electron in the disc will feel a radial electric field. The circuit is completed by having an outside wire from the axle to the outside of the disc.

Conceptually, the simplest generator is a single loop of area A rotated at constant angular speed ω in a constant uniform magnetic field $\vec{B}$, the axis of the loop being perpendicular to $\vec{B}$. The magnetic flux through the loop, as previously discussed, is $\Phi_B(t) = AB\cos\omega t$ so the induced emf is



$\mathcal{E}(t) = AB\omega\sin\omega t$. This is an AC (alternating current) generator. If the two ends of the loop are connected via slip rings to an external resistive circuit, current will flow. This current will of course give the loop a dipole like magnetic field, which from Lenz' law will be such as oppose the imposed rotation. This means that once the circuit is complete, whatever is maintaining the rotation will have to work harder, obvious from energy conservation, since the circuit is now generating heat in its wires, and possibly useful work as well. It's worth checking with a small generator, such as for lights on a bicycle, how much harder it is to turn when connected to a circuit.

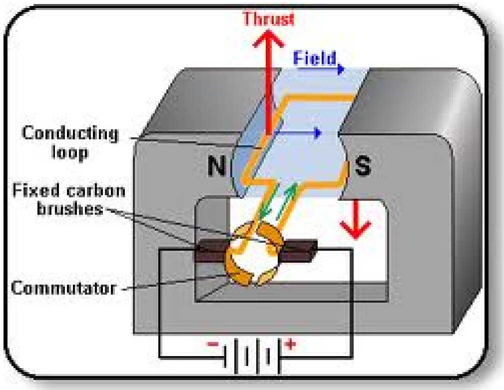
Eddy Currents



In fact, any time magnetic field strength is changing, circling electric fields are generated, and if these fields are in a conducting medium, currents will arise creating magnetic fields partially compensating for the changes taking place in the original magnetic field. These currents are called eddy currents—they are reminiscent of the circling eddies in the wake of a boat, since they are generated by the nonconservative circling electric fields. Notice the opposite directions of the eddy currents—on entering the magnetic field, the currents oppose the field increase, on leaving, they attempt to maintain the field strength. Induction cookers use eddy currents.

21 Magnetic Induction III

Electric Motors



One common electric motor design is just the "loop in a magnetic field" generator we already discussed, but now run backwards—by which we mean power is supplied from a battery, say, to the loop which then becomes a dipole-like electromagnet. The loop is between the poles of a permanent magnet so it turns, but as it reaches the limit a commutator reverses its current supply, so attraction suddenly becomes repulsion and since it's rotating it continues around, whereupon the sequence repeats.

Back emf

As the loop rotates in the magnetic field*, that rotation will induce an emf in the loop opposing the motion*—in other words, opposing the driving emf!
This is called *back emf*, and is proportional to speed.

When a motor is first connected, it is not turning and Ohm's law gives $V_0 = IR,$ where V_0 is the voltage of the supply, and R the resistance of the armature (meaning the loop or coil). Heat production inside the motor is $I^2R.$

When the motor is running under load, there is a back emf V_{back}, and now $V_0 - V_{\text{back}} = IR.$

Heat production in the motor is now I^2R: which can be *much less* than it was initially.

If a blender is mechanically overloaded so the motor turns slowly, back emf is small, the current is higher than designed for, high heat production for some time may cause burnout.

Back emf problem:

A motor has an armature resistance of 4Ω.

It draws 10A from a 120-V line when running at its design speed of 1000 rpm.

If a load slows it to 250 rpm, what is the current in the armature?

Counter Torque

A generator is essentially a loop rotating in a magnetic field.

If the generator is connected to an outside circuit, the induced emf will cause a current to flow: that's the point of the generator!

But the current carrying wire moving through the field will feel Lenz-type forces opposing its motion: called the "counter torque".

So to produce a current through the external circuit work must be done. Obviously.

An example from the book: #34.

A conducting rod, mass *m*, resistance *R*, length ℓ, rests on two frictionless and resistanceless parallel rods, in a perpendicular magnetic field *B*. At *t* = 0, a source of emf is supplied to the rails. How does the rod move, if case (a) the source maintains constant current *I*, and case (b) the source supplies constant emf $\mathcal{E}$?

Case (a): in a magnetic field B, the force on a length of wire carrying a current is $\vec{F} = I\vec{\ell} \times \vec{B}$, so if *B* is inwards, and the current is flowing downwards (top rail positive) the rod will feel a constant force $I\ell B$ to the right, and hence accelerate at a uniform rate.

Case (b): the rod will still feel accelerating force $I\ell B$, but now the current will vary because the motion of the rod through the field generates an emf in the rod, of magnitude $\mathcal{E} = -d\Phi_B / dt = -v\ell B$. You can check that this emf is opposing the driving emf: the rod is moving to the right, so the force on a charge *q* in the rod is $q\vec{v} \times \vec{B}$, upward for a positive charge, opposing the external voltage. But this is also just Lenz' law: the induced emf is such as to oppose the motion.

Now, the current I is given by $I = (\mathcal{E} - v\ell B)/R$, and the force accelerating the rod is $I\ell B$, so

$$m\frac{dv}{dt} = I\ell B = \frac{(\mathcal{E} - v\ell B)\ell B}{R}$$

From which

$$\frac{dv}{\mathcal{E} - v\ell B} = \frac{\ell B}{mR}dt$$

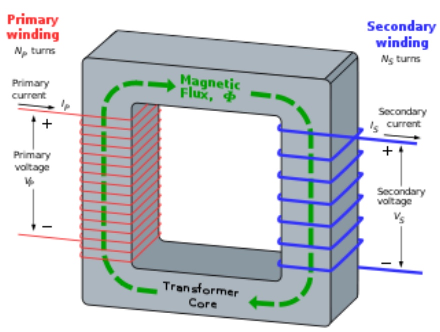
You may recognize the form of this equation: it's the same as the one we found for charging a capacitor. The left-hand side integrates to a log, the right hand side is trivial, a constant times *t*. We'll leave that as an exercise: the result is just like charging a capacitor, initially the rod accelerates at a constant rate, but there's a limiting speed, as *v* approaches $\mathcal{E} / \ell B$ the left-hand side blows up, so to match it longer and longer times elapse. The speed approaches this value but never quite reaches it.

What's happening to the current in the rod at this stage? It's getting smaller and smaller, remember $I = (\mathcal{E} - v\ell B)/R$.

This is an example of back emf: if an external voltage is used to supply a current to a moveable wire, which is in a magnetic field and moves because of the Lorentz force, then the magnetic field induces an emf in the conductor opposing the supplied voltage. Again, this is just Lenz' law.

Back emf plays an important role in electric motors, and can be a big fraction of the applied emf. Of course, there is no back emf in a jammed motor, and this is why a jammed motor will likely burn out—the coils are only designed to take the large current from the unopposed external emf for a short time, the design assumes there will be back emf when the motor is in use, limiting the current.

Transformers



The big reason power is transmitted as ac rather than dc is that it's very easy to change the voltage using transformers for ac, and much less power is lost in transmission lines that run at high voltage. A transformer has two solenoid-type coils: the input power goes into the primary, producing a constantly changing magnetic field, this changing flux induces oscillating emf in the secondary coil. To maximize the effect, soft iron or some similar alloy is used to guide all the flux from the primary through the secondary. This iron is laminated—it's in thin sheets separated by very thin layers of insulation, to minimize eddy current buildup.

Neglecting energy loss from eddy currents and Ohmic loss in the coils themselves, the back emf in the primary must be balancing the input voltage, and this back emf equals $N_P d\Phi_B / dt$, where Φ_B is the *total* magnetic flux through the coil, this flux coming from both the primary coil *and* the secondary coil if that is part of a circuit. What is the output voltage? The same rate of change of flux takes place in every turn of the secondary (output) coil also, so

$$\frac{V_S}{V_P} = \frac{N_S}{N_P}.$$

The ratio of voltages is just the ratio of the numbers of turns!

The Betatron

This is a very clever device for accelerating electrons to very high speeds (close to the speed of light). It uses a magnetic field to make the electrons go in circles of radius r, as earlier (for the cyclotron) we have

$$mv^2 / r = evB(r), \quad mv = reB(r)$$

where we have a perpendicular magnetic field with strength a function of r, arranged by appropriately shaped poles of an electromagnet.

Now, we increase the magnetic field: this generates a momentary circling electric field,

$$\oint \vec{E} \cdot d\vec{\ell} = -d\Phi_B / dt$$

and for a circular path $2\pi rE = \pi r^2 d\overline{B} / dt$, or $E = (r/2) d\overline{B} / dt$, where $\overline{B}$ is the average magnetic field inside the circle.

In the betatron, the magnetic field is increases in such a way that the electrons continue to circle at the same radius, but are speeded up by the electric field generated by the increasing magnetic field.

How is that possible?

$$\frac{d(mv)}{dt} = E$$

and if r is constant,

$$\frac{d(mv)}{dt} = re\frac{dB(r)}{dt} = eE = \frac{er}{2}\frac{d\overline{B}}{dt}$$

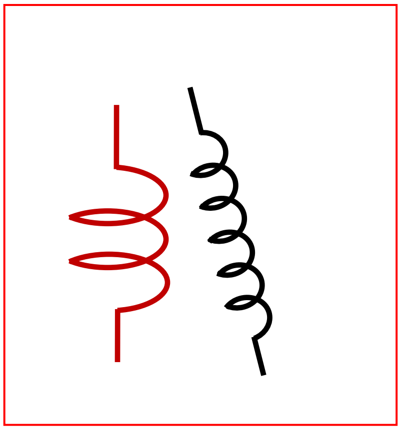
so provided $B(r) = \overline{B}/2$, the electrons will continue to orbit at the same radius! The pole pieces can be designed so this is true at the design radius.

The betatron, unlike the simple cyclotron, can accelerate electrons to relativistic speeds, where they have greatly increased mass.

22 Mutual Inductance

Definition, and an Important Symmetry

We've already discussed how an ac transformer operates: the alternating current in the first coil generates an oscillating magnetic field, the soft iron guides all the magnetic flux through a secondary coil, where the changing flux generates an ac emf, which can be tapped off as a power source. In operation, the magnetic flux from both coils threads both coils, and a complete analysis requires solving coupled equations. But since both coils experience the same changing flux, it follows immediately that the emfs thereby generated are in the ratios of the number of turns, and since the emf in the primary is just balancing the supplied external emf, voltage in secondary/voltage across primary = N_2/N_1. Neglecting the tiny losses from Joule heating (including eddy currents), the power absorbed in the primary must be the power supplied by the secondary, so the ratios of the currents will be the *inverse* of the voltage ratios.



More generally, if two coils 1, 2 are in proximity, a current through one will cause some magnetic flux to thread through the other, so a changing current in one will induce an emf in the other. From the Biot-Savart Law, the magnetic field from a current is linear in the current (directly proportional to it) so from Faraday's law, the induced emf will be linear in the rate of change of the current I_1 in coil 1.

The coefficient of proportionality is called the **mutual inductance**, and is denoted by *M*:

$$\mathcal{E}_2 = -M_{21}\frac{dI_1}{dt}.$$

Putting in the minus sign is standard practice—as usual, the direction of the emf and consequent current should be found using Lenz' law.

The mutual inductance can also be expressed purely in terms of the magnetic flux linkage:

it's just the total magnetic flux through coil 2 when there is unit current in coil 1.

Writing this total flux as $\Phi_{I_1=1}$, for current I_1 in coil 1 the total flux through coil 2 is $\Phi_{I_1=1}I_1$, and for a changing current in coil 1 the induced emf in coil 2 is given by

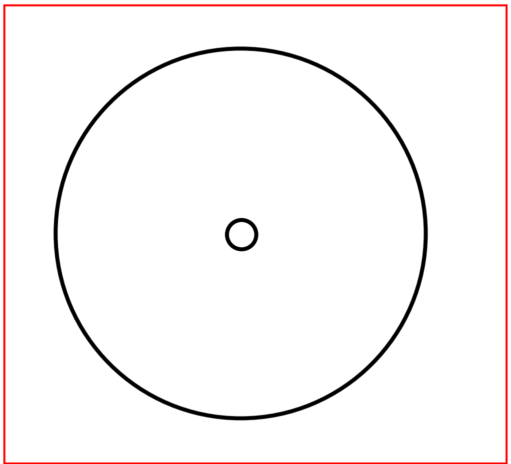
$$\mathcal{E}_2 = -\frac{d\Phi}{dt} = -\Phi_{I=1}\frac{dI_1}{dt}$$

It turns out that for two coils, or indeed for any two current-carrying conductors, the **mutual inductance is symmetric**:

$$M_{12} = M_{21}.$$

This is by no means obvious! It cannot be proved using simple arguments—it is necessary to use vector calculus. Recall that the electrostatic potential could be written in terms of a potential. It turns out that the magnetic field can be written as the curl of a vector potential, and this formulation plays an essential role in the proof of symmetry of mutual inductances.

This symmetry can be handy! Quite often, it's easy to evaluate the inductance one way, but not the other. For example, consider trying both approaches when coil 1 is a long solenoid going through a single loop, this loop being coil 2.



Another example: consider a small circular loop at the center of a large circular loop, both in the same plane. Putting a current I in the big loop gives a field on the axis at the center of $\mu_0 I / 2R$, so if the small loop has *r* << *R*, we have

$$M_{12} = \mu_0 \pi r^2 / 2R.$$

Notice this goes down as *R* increases. But consider M_{21}: the little loop has a dipole-like field, certainly at large distances, like *R*. This field goes through the large loop.

But how can the amount of *this* field going through the big loop go *down* as the size of the big loop increases? (It must, since $M_{12} = M_{21}$, and M_{12} certainly goes down.)

Because if the big loop is infinitely large, the total flux through the big loop from a current in the little loop is zero! The magnetic flux goes up in the middle, through the little loop, then back down again outside the little loop (remember the magnetic field lines just circulate around). For a large but finite big loop, the net flux through the big loop from a current in the little loop is just the negative of the flux passing outside the big loop.

Self-Inductance: Energy Stored in the Magnetic Field

We've already discussed back emf, a changing current in a coil generates an emf opposing the change in current. The ratio of the induced emf to the rate of change of current is called **self inductance**, and written *L*,

$$\mathcal{E} = -L dI / dt.$$

(As usual, the minus sign is there to remind us that the emf is opposing the current increase—it is necessary to use Lenz' law, or equivalently energy considerations, to find emf direction in any particular case.)

As increasing current is supplied to a coil, this back emf forces the external power supply to do work against it at a rate $P = I\mathcal{E} = LIdI / dt.$

Therefore the total work done is increasing the current from zero to *I* is $L\int IdI = \frac{1}{2}LI^2.$

This is directly analogous to charging a capacitor: in that case, it took more work to add more charge, and we saw the energy was stored in the electric field, with density $\frac{1}{2}\varepsilon_0 E^2$. For the inductance, the energy is similarly stored in the **magnetic** field. We can check that for a solenoid: for *N* turns and length ℓ, the field inside the solenoid $B = \mu_0 NI / \ell$.

The self-inductance is the total flux through the *N* turns **for unit current**, take the cross-sectional area to be *A*, this is

$$L = NAB = \mu_0 N^2 A / \ell.$$

Writing $I = \ell B / \mu_0 N$, we find

$$\tfrac{1}{2}LI^2 = \frac{1}{2}\frac{\mu_0 N^2 A}{\ell}\left(\frac{\ell B}{\mu_0 N}\right)^2 = \tfrac{1}{2}A\ell B^2/\mu_0$$

That is, the magnetic energy density is $B^2/2\mu_0$. Recalling that $\mu_0 = 4\pi\times 10^{-7}$, a cubic meter at one tesla contains megajoules, highly relevant for designers of big particle accelerators, where the current is flowing in superconductor. If these are accidently heated, they become normal resistors, with consequent sudden—explosive—loss of field. This has happened.

We can use this energy density approach to find the inductance of a coaxial cable quite easily:

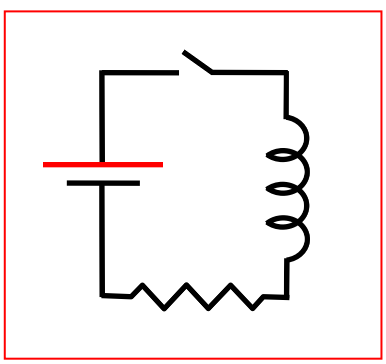
Between r_1, r_2 the magnetic field is $B = \mu_0 I/2\pi r$, so the energy density per unit length is

$$\frac{1}{2\mu_0}\int_{r_1}^{r_2} B^2 2\pi r dr = \frac{1}{2\mu_0}\int_{r_1}^{r_2}\left(\frac{\mu_0 I}{2\pi r}\right)^2 2\pi r dr = \frac{1}{2}\left(\frac{\mu_0}{2\pi}\right)\ln\frac{r_2}{r_1} I^2$$

from which $L = \frac{\mu_0}{2\pi}\ln\frac{r_2}{r_1}$ per meter. Notice that letting r_2 go to infinity, we conclude that a single infinitely long wire of radius r_1 has infinite magnetic energy density *per unit length*! Of course, there is no such thing as an infinitely long wire, and the divergence is very slow. It is also apparent that the self-inductance of, for example, a single loop of wire must depend on the radius of the *wire* as well as that of the loop.

23 *LR*, *LC* and *LRC* Circuits

LR Circuits



A battery is connected to *L*, *R* in series: $V_0 = IR + L\dfrac{dI}{dt}$.

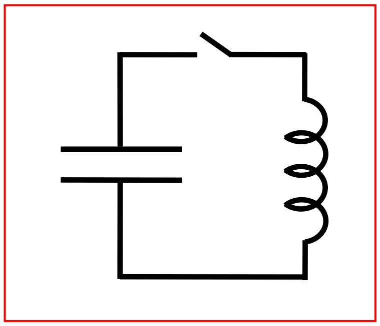
(*Quick sign check*: on first connecting the battery, *L* will oppose the current increasing from zero—so for positive dI/dt, the inductance is working against the battery.)

The math here is very similar to the capacitor:

$$\frac{dI}{V_0 - IR} = \frac{dt}{L}$$

The integration is routine, and $I = \frac{V_0}{R}\left(1 - e^{-t/\tau}\right)$ with $\tau = L/R$.

LC Circuits: Oscillations



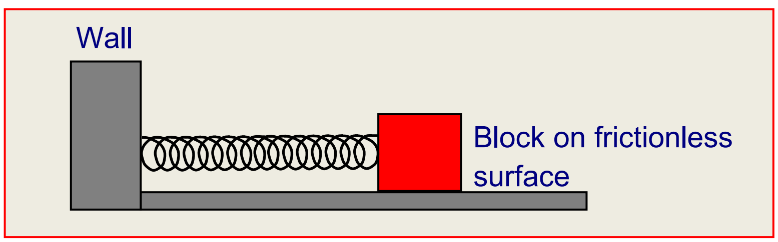
We first assume there's zero resistance, just *C* and *L* in series, with an open switch. We charge *C*, then close the switch. Charge will begin to flow through the inductance from one plate of *C* to the other. The current $I = -dQ/dt$ will build up, but its *rate of increase will depend on the inductance,* the back emf LdI/dt balancing the voltage Q/C of the capacitor.

That is, $\frac{Q}{C} = L\frac{dI}{dt} = -L\frac{d^2Q}{dt^2}$ or $L\frac{d^2Q}{dt^2} = -\frac{Q}{C}$.

Now *compare this charge equation with the equation for displacement of a mass on a spring*: suppose the mass can slide on a smooth table, the spring being attached to a wall. The equation is:

$$m\frac{d^2x}{dt^2} = -kx.$$

Here *m* is the mass, *k* the spring constant, and *x* the linear displacement from the spring's relaxed length position. It's clear that *these two equations are mathematically identical!*



It's worth seeing what corresponds to what. The inductance *L* corresponds to the mass *m*. These are both "inertial" terms: once the mass is moving, its mass or inertia keeps it going, so if it's pulled to one side then let go, the spring pulls it back, accelerating it, but even when the spring is back to its natural length the mass keeps moving and takes the mass the same distance the other way before stopping. For the *LC* circuit, the current is analogous to the velocity of the mass, and the inductance opposes change in current just as mass opposes change in velocity: when the initially charged capacitor has completely discharged, the inductance keeps the current going until the capacitor has its initial charge reversed.

We can see from the equations that the spring constant *k* corresponds to the inverse of the capacitance, 1/*C*, so a bigger capacitance is like a softer spring. A larger capacitance can absorb charge more easily, just as a softer spring can be more easily stretched.

For the mass on the spring on a frictionless surface, pulled aside by x_0 and let go, $x = x_0 \cos \omega t$, here $\omega = \sqrt{k/m}$.

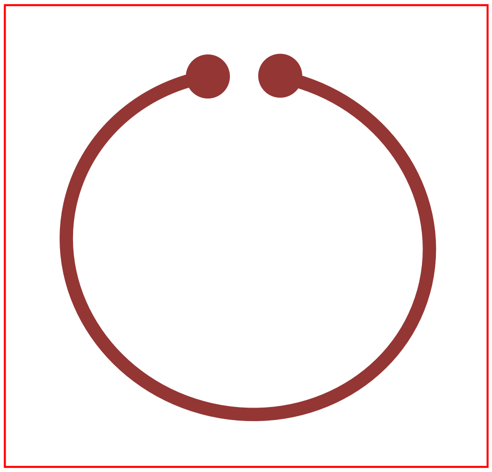
Similarly, if at $t = 0$ there is zero current and charge Q_0 on the capacitor, $Q = Q_0 \cos \omega t$, where $\omega = 1/\sqrt{LC}$.

The trading of *potential energy* stored in the spring and *kinetic energy* in the moving mass, with overall energy conserved at all times, is mirrored here in the trading of *electric field energy* in the capacitor and *magnetic field energy* in the inductance.

To check *conservation of energy* here, note that $Q = Q_0 \cos \omega t$ means that $I = -dQ/dt = Q_0 \omega \sin \omega t$, and total energy

$$U_{\text{total}} = \frac{Q^2}{2C} + \frac{LI^2}{2} = \frac{Q_0^2 \cos^2 \omega t}{2C} + \frac{LQ_0^2 \omega^2 \sin^2 \omega t}{2} = \frac{Q_0^2}{2C}$$

using $\omega^2 = 1/LC$.



It's also worth mentioning here that this *L*, *C* circuit can be very simple: for example, a ring with a gap:

The two balls are the capacitor: begin with one of them positive, the other equally negative. The ring is of course an inductor: a current going around it creates a magnetic field.

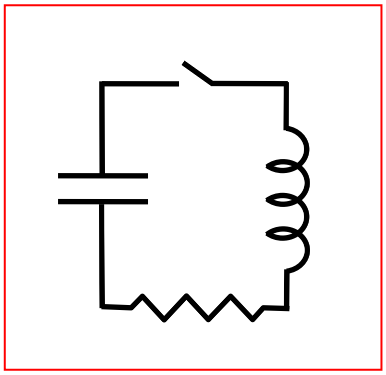
Initially, then, we have an electric dipole, with the corresponding field. But this is an oscillating field. The magnetic field from the current also oscillates, and we know that creates another electric field. The picture is not yet complete—there is one crucial element missing—but we can begin to see how an oscillating circuit might emit electromagnetic waves.

LRC Circuit

How does this picture change if we include resistance? Obviously, the same way, essentially, the oscillating mass on a spring changes if we include air or fluid resistance: the oscillations are damped.

More formally, we must add a term *IR* to the voltage sum:

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{Q}{C} = 0$$



This equation for *Q* is *identical* to the equation of motion for *x* for a *damped* mass on a spring: imagine a mass in molasses. *L* is the mass, 1/*C* is the spring constant *k* and *R* is the damping term *b*: see that earlier work.

This means that the behavior of *Q* as a function of time is exactly the same as the behavior of *x* in the damped oscillator. In particular, there are two different regimes: lightly damped, where the sign of charge on the capacitor oscillates back and forth, gradually dying away, and heavily damped, where the charge slowly drains through the system, decreasing exponentially, with no oscillation. The boundary between the two is critical damping, and occurs when the damping term $R^2 = 4L/C$.

It also follows from the exact analogy with the damped mechanical oscillator that introducing damping (resistance) lowers the frequency of the oscillations, but for small damping this is a very small effect. However, as the damping approaches the critical value $R^2 = 4L/C$, the frequency goes to zero.

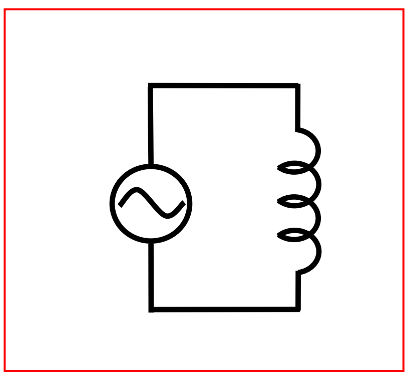
home

24 Circuits with AC Source

Pure Resistance *R*

If we put an alternating voltage $V = V_0 \cos\omega t$ across a *resistor* R, from Ohm's law the current will be $I = I_0 \cos\omega t$ and $V_0 = I_0 R$. The power dissipation in the resistor $\overline{P} = I_0 V_0 \overline{\cos^2 \omega t} = \frac{1}{2} I_0 V_0 = V_{\text{rms}}^2 / R$.

Pure Inductance *L*



If we put an alternating voltage across an *inductor L*, having zero resistance *R*, this external voltage must be exactly matched by the back emf generated.

That is, $V = LdI/dt$. So if $V = V_0 \cos\omega t$, we must have a current $I = I_0 \sin\omega t$, with $V_0 = \omega L I_0$.

This looks a lot like Ohm's law, so by analogy we write

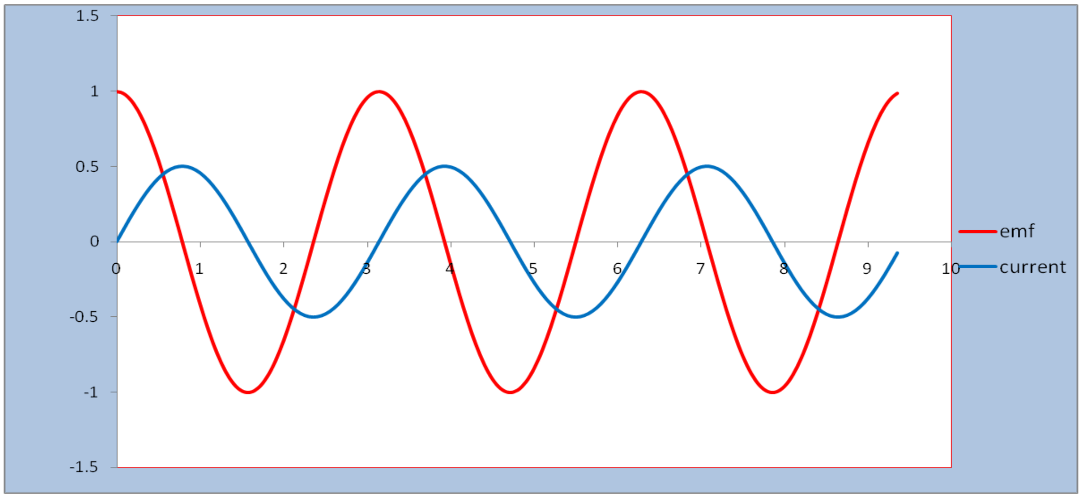
$$V_0 = I_0 X_L$$

and call X_L the inductive *reactance*.

But there's a big difference from a resistance: *no power* is dissipated by a pure inductance: $\overline{P} = I_0 V_0 \overline{\sin\omega t \cos\omega t} = \frac{1}{2} I_0 V_0 \overline{\sin 2\omega t} = 0$.

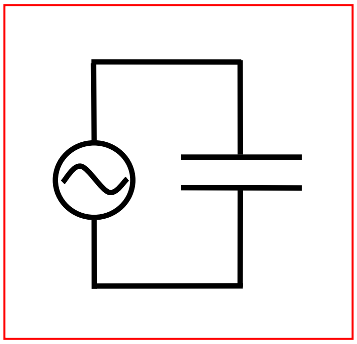
It's worth plotting voltage and current as functions of time on the same graph. For $V = V_0 \cos\omega t$, at $t = 0$ the voltage is at its maximum positive value, so the current is increasing at a maximum rate. One-quarter of a cycle later ($\omega t = \pi / 2$) the external applied voltage is zero, so at that instant the current is not changing—in fact, the current has reached its maximum positive value. So we see the current reaches its maximum one quarter of a cycle (or $\pi/2$ or 90°) after the maximum voltage: we say the current lags behind the voltage by 90°.

Note: the graph below, and the subsequent ones on AC circuits, were generated by an Excel spreadsheet available for download from my 2415 Home Page. Try it—it's a good way to explore these circuits!



Current and emf in a purely inductive AC circuit: emf leads current by a quarter cycle (ELI).

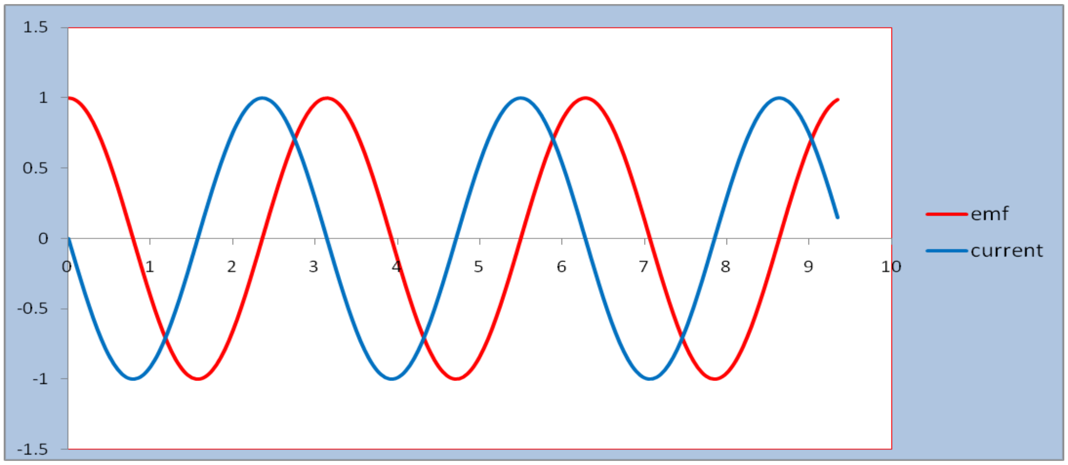
Pure Capacitance *C*



Putting $V = V_0 \cos \omega t$ across a capacitor, again with no resistance (or inductance) in the circuit, at $t = 0$ the voltage is at its maximum positive value, which means the capacitance contains its maximum positive charge in the cycle. At that moment, the current must be zero: the capacitance has charged up fully, and is about to begin discharging as it follows the cycle.

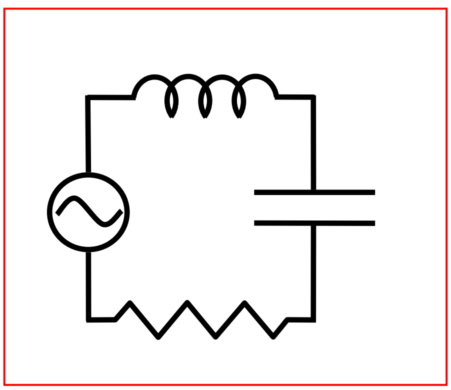
The equations are $V = V_0 \cos \omega t = Q / C$ and $I = dQ / dt = -\omega C V_0 \sin \omega t = -I_0 \sin \omega t$.

We write $V_0 = I_0 X_C$ where X_C is the *capacitive reactance* $X_C = 1 / \omega C$.



Current and emf in a purely capacitative AC circuit: current leads emf by a quarter cycle (ICE).

Driven LRC Circuit



The equation describing charge variation as a function of time is:

$$L\frac{d^2Q}{dt^2}+R\frac{dQ}{dt}+\frac{Q}{C}=V_0\cos\omega t$$

where we've arbitrarily fixed time zero to correspond to a maximum of the driving field. Obviously, if the AC field is turned on at time zero, there is some transient behavior before it settles down. That might be important in some situations, but we're only going to analyze the "steady state" situation, the rhythm the system settles into after the AC has been on for many cycles.

The equation can be solved (most simply using complex numbers); we'll skip the derivation and present the solution for the current. As one would expect, it has the same period of variation as the driving emf, but is not in general in phase. Explore the solutions using the spreadsheet!

The current is $I=I_0\cos\left(\omega t-\varphi\right)$ where

$$I_0=\frac{V_0}{\sqrt{\left(X_L-X_C\right)^2+R^2}}=\frac{V_0}{\sqrt{\left(\omega L-\left(1/\omega C\right)\right)^2+R^2}},\quad \varphi=\tan^{-1}\left(\frac{X_L-X_C}{R}\right).$$

(You can of course check that this is a solution by plugging it into the equation.)

The form of φ tells us immediately that for a purely inductive circuit, the emf leads the current (ELI), for a purely capacitative circuit, current leads emf (ICE).

Resonance

Something surprising happens if $\omega L = 1/\omega C$—there is no phase lag, and the current amplitude is given by Ohm's law for the resistor alone! Notice this is the same value of ω at which a pure *LC* circuit oscillates. This is resonance: the driving frequency coincides with the natural frequency of the circuit, and the amplitude of the oscillations is limited only by the damping—the resistance.

Another way to understand this is to remember that the current through all three elements (*L*, *C*, *R*) of the circuit has the same phase. From Ohm's law, the voltage drop across the resistor oscillates exactly in phase with the current. The voltage change across the inductance has phase 90° *ahead* of the current's phase, that across the capacitor has phase 90° *behind* the current. This means that the voltage changes across the inductance and the capacitor are 180° out of phase—meaning they are opposite, so if the amplitudes are equal, they'll exactly cancel! (Check this on the spreadsheet.)

Power Factor in an AC Circuit

As usual, $\overline{P} = \overline{VI}$—in this case, using a standard trig identity, and $\overline{\cos^2 \omega t} = \frac{1}{2}$, $\overline{\cos \omega t \sin \omega t} = 0$, we have

$$\begin{aligned}
\overline{P} = \overline{VI} &= V_0 I_0 \overline{\cos \omega t \cos\left(\omega t - \varphi\right)} \\
&= V_0 I_0 \overline{\cos \omega t \left(\cos \omega t \cos \varphi + \sin \omega t \sin \varphi\right)} \\
&= \tfrac{1}{2} V_0 I_0 \cos \varphi = V_{\text{rms}} I_{\text{rms}} \cos \varphi.
\end{aligned}$$