UCD: PHYSICS 9C - ELECTRICITY AND MAGNETISM

Tom Weideman
University of California, Davis

LibreTexts™

University of California, Davis

UCD: Physics 9C - Electricity and Magnetism

Tom Weideman

The LibreTexts libraries are Powered by MindTouch® and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739. Unless otherwise noted, LibreTexts content is licensed by CC BY-NC-SA 3.0.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

This text was compiled on 12/10/2020

TABLE OF CONTENTS

Physics 9C introduces general principles and analytical methods used in electricity & magnetism for physical science and engineering majors.

1: ELECTROSTATIC FIELDS

We begin with a study of electric fields due to static charge distributions.

2: ELECTROSTATIC ENERGY

Now that we have described electrostatic phenomena in terms of vector fields that relate to the forces on charges, we now turn to scalar fields, which relate to the potential energies of charges.

3: DIRECT CURRENT CIRCUITS

We finally discard our requirement that charge remains static, and examine the effects associated with moving charge.

4: MAGNETISM

The phenomena of electricity and magnetism turn out to be two sides of the same coin, but we follow the historical path of examining them as if they are separate forces. We therefore set aside electricity and dive into magnetism.

5: ELECTROMAGNETISM

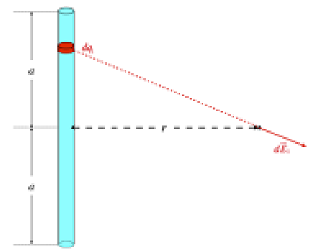
The time has come to put our two disparate forces together. We know that they have electric charge in common as a source of field and a recipient of force, but here we will see that they have much more in common than this.

BACK MATTER

CHAPTER OVERVIEW

1: ELECTROSTATIC FIELDS

We begin with a study of electric fields due to static charge distributions.



1.1: CHARGES AND STATIC ELECTRIC FORCES

The electric force is the fundamental force behind nearly every macroscopic force we studied in classical mechanics.

1.2: ELECTRIC FIELD

Originally intended as merely an "explanation" of action-at-a-distance, the idea of an electric field turns out to be a much more useful model than one might expect.

1.3: COMPUTING ELECTRIC FIELDS FOR KNOWN CHARGE DISTRIBUTIONS

When we encounter electric charges in the real world, they appear in very great numbers. This allows us to treat them as approximately a continuous distribution, making integral calculus a powerful tool for field calculation.

1.4: DIPOLES

Particles we encounter (such as atoms and molecules) rarely are electrically charged, as they tend to attract and bond with other particles that are oppositely-charged. But these neutrally-charged particles are still affected by electric fields, thanks to their component charges being ever-so-slightly separated.

1.5: CONDUCTORS

It is useful to model materials as one of two types: Those that allow charges to flow freely through them, and those that do not. Here we will examine properties of the former.

1.6: GAUSS'S LAW

The only link we have seen between charge and electric field is Coulomb's law, coupled with the principle of superposition. It turns out that these two quantities have a much deeper relationship, which can be exploited to solve problems in a manner easier than what we have seen so far.

1.7: USING GAUSS'S LAW

Gauss's law has a number of practical uses, such as computing electric fields for highly-symmetric situations, and dealing with conducting shells.

1.8: METHOD OF IMAGES

We develop a trick that allows us to discuss forces and fields that result from bringing free charges near flat plane conductors.

1.1: Charges and Static Electric Forces

Fundamental Forces

In Physics 9A, we studied forces. There were many such forces to discuss – gravity, tension, contact (normal), friction, etc. There was one major distinction between gravity and the others – gravity acts at a distance, while all the others required physical contact. In physics we call the different types of forces that act at a distance *fundamental forces*, inasmuch as all the other forces are built from them. That's right, ultimately *all* forces act at a distance – "contact" has no meaning in the microscopic realm of tiny particles. The forces just mentioned, such as tension and friction, are at their core really electrical in nature. There isn't actually any "contact" made – it's just that the separation is so small that we can't see the forces between individual particles with the naked eye, though we *can* see the aggregate of the forces between many such particles. For example, the force that causes a statically-charged balloon to cling to a wall, or freshly-combed hair to fly apart are forces that act at a distance and can be witnessed with the naked eye.

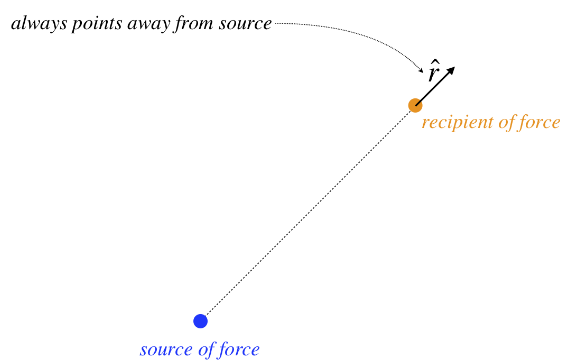
Given that we are seeing this force act at a distance, we might be inclined to compare it with the only other force we know of that behaves in this manner: gravity. Let's have a look back at Newton's law of universal gravitation…

$$\overrightarrow{F}_{m_1 \ on \ m_2} = \frac{Gm_1m_2}{r^2}(-\hat{r}) \tag{1.1.1}$$

Let's break this down, piece-by-piece:

- G – This is just a constant – it ensures that the units come out to what we are accustomed to.
- m_1 and m_2 – These are the masses of the two gravitating bodies (technically they are either point masses or are spherically-symmetric, but this is quibbling). Mass is both the *source* of the force and the reason an object reacts to the force. That is, m_1 creates the gravity force that m_2 feels. When the gravity source m_1 is doubled, the force felt by m_2 is doubled. When the gravity source m_1 is left unchanged but the mass of the object m_2 is doubled, the force on that object is again doubled.
- r – This is the separation of the two point masses (or the centers of the two spherical masses). We see that the force gets *weaker* as the separation increases, according to an *inverse-square law*. If the separation is doubled, the force gets weaker by a factor of $\frac{1}{4}$. If the separation is tripled, the force gets weaker by a factor of $\frac{1}{9}$, and so on.
- $-\hat{r}$ – This is the unit vector that indicates the direction of this force. It is defined as pointing *away from the source of the force*. The minus sign indicates the opposite direction, which is toward the source. The force is acting on the other object, so since the direction of the force on it is toward the gravity source, gravity is said to be an *attractive force*.

Figure 1.1.1 – Definition of $\hat{r}$



So given what we know about gravitation, we can explore the action-at-a-distance force we observe with static electricity.

Coulomb's Law

First of all, we know that the static electric force is distinctly different from gravity on two different counts. First, the if we have two neutrally-charged particles, we don't see this amount of force, so it can't be their masses that are responsible for it. And second, besides an attractive force, we also witness a *repulsive* force, which we do not see for gravity.

A little playing around with the attractive and repulsive forces shows us that there must be two different types of quantities responsible for the force. Gravity had only one type – mass – but this new *electric force* must have two different types of "mass," which we refer to as *electric charge*. The units for this quantity are named after the fellow who did the first detailed exploration of the force, *coulombs* (C). When we play around with this force, we find that when two of the same kind of charge are brought together, the force is repulsive, while when we bring different types of charge together, the force is attractive. We summarize this phenomenon with the commonly-used maxim:

unlike charges attract, while like charges repel

For now, we can call these two types of charge whatever we like – black and white, left and right, dumb and dumber – anything that distinguishes them from each other. We'll come back to a convenient description of the two types shortly.

Next we need to explore the other elements of the force, as we did for gravity. Here are the things we find:

- By holding one charge fixed and varying the other, we can determine the relationship between the amount of charge causing or experiencing the force and the strength of the force. We find that the strength of the force (like gravity) is proportional to both the amount of charge *causing* the force and the amount of charge *feeling* the force. So for gravitation we had m_1 and m_2, and now for electric force we have q_1 and q_2 (q is the traditional variable used to represent electric charge). The force between two point charges is proportional to the product q_1q_2.
- We can also test the dependence of the force between two charges on the separation of those charges. Remarkably, we find that when the separation is doubled, the force goes down by a factor of $\frac{1}{4}$, and when it is tripled it goes down by a factor of $\frac{1}{9}$, exactly as it does for gravitation! So the electric force also obeys an inverse-square relation.

Putting these two things together with a constant (similar to G, but with different units) that exists to make our units work out correctly, we end up with a magnitude of the electric force that obeys:

$$\left|\overrightarrow{F}_{q_1\;on\;q_2}\right| = \frac{kq_1q_2}{r^2}, \quad where: \;\; k \equiv 9.0\times10^9\,\frac{Nm^2}{C^2} \tag{1.1.2}$$

All that is missing is our vector direction, which we covered nicely with the $-\hat{r}$ in the case of gravity. But what do we do here, append a little explanation, "repulsive if both charges are the same type, attractive if they are not?" That is not very satisfying mathematically, and it turns out there is a nice, elegant solution: Call one of the types of charge "positive," and the other "negative," and write the force vector as follows:

$$\overrightarrow{F}_{q_1\;on\;q_2} = \frac{kq_1q_2}{r^2}\hat{r} \tag{1.1.3}$$

The $\hat{r}$ means the same as it did before – it points away from the charge causing the force. But notice what our mathematical definition of positive and negative charges accomplishes: If both charges are the same type, then their product is positive, and the force points in the direction of $\hat{r}$, which is repulsive (object is pushed away from the source). But if the charges have opposite signs, then their product is negative, and the force direction is $-\hat{r}$, resulting in an attractive force. The above force law is called *Coulomb's law*.

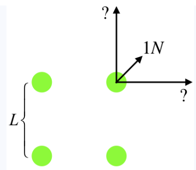
As with any forces, when there is more than one electrical force on a charge, the total force is computed by adding the individual forces like vectors. We call this principle *superposition*.

Example 1.1.1

Four identical charges are located at the corners of a square. The force exerted between two charges at opposite corners is $1N$. Find the magnitude of the net force on one of the charges.

Solution

First of all, with the charges being identical, the force between any pair of charges is repulsive. In the diagram below, we have labeled the length of each side of the square, and drawn in the three forces on one of the charges, one of which is due to the diagonal charge, and the other two are unknown, but they are equal in magnitude to each other, and are greater than $1.00N$ because the charges are identical and the adjacent charges are closer than the diagonal charge.



Writing the coulomb force between the diagonal charges in terms of the charges (which we'll call Q), and the separation (which across a square of side L is $\sqrt{2}L$), we can get the magnitude of the force exerted between adjacent charges, which are separated by a distance L:

$$1.00N = \frac{kQ^2}{\left(\sqrt{2}L\right)^2} \quad \Rightarrow \quad \frac{kQ^2}{L^2} = 2.00N$$

Now we superpose the three forces (i.e. add them like vectors) to get the magnitude of the net force, Fortunately this is easy, as the direction of the sum of the vertical and and horizontal force vectors points in the same direction as the diagonal force vector:

$$F_{net} = \sqrt{(2N)^2 + (2N)^2} + 1N = 3.83N$$

Properties of Electric Charge

There are just a few things we need to say about electric charge before we move on.

1. There exists a minimum "fundamental" unit of charge, below which charge cannot be subdivided. This amount is a very small number of coulombs:

$$e \equiv 1.6 \times 10^{-19} C$$

This is the amount of charge that resides on the tiny particles we have heard about that reside within atoms: protons (+ charges), and electrons (– charges). We have Benjamin Franklin to "thank" for these sign conventions (we'll see that this choice is a bit of a pain). Despite the fact that charges come in integer numbers, we will treat them as continuous values.

2. Electric charge is *conserved*. We know from our study of energy what this means – it can neither be created nor destroyed. It doesn't mean that the particles on which that charge resides cannot change into other types of particles (e.g. neutron → proton + electron + antineutrinos), just that when the particles do change, the new particles' charges have to add up to the same charge the original particle started with.
3. When we deal with charges in the real world, they are usually residing on some substance. We can very crudely divide materials into two types of substances (actually there are many variations in-between, but for now this distinction will do):
    - *conductors (think "metals") – charges are free to move around on (and through) these substances*
    - *insulators (think "non-metals") – charges are locked into place and cannot move around*

As innocuous as these distinctions may seem, we will see that they lead to some profound properties.

1.2: Electric Field

Coulomb Field

While one can describe the details of forces between charges mathematically, it still is very unsatisfying – *how* do the charges affect each other from a distance? This question troubled physicists for a long time, and the "solution" (really it is just a model that works) is quite ingenious. It goes like this:

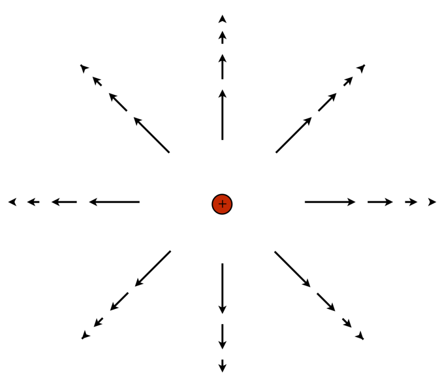
The source of the electric (or for that matter, gravitational) force doesn't know anything about the existence of another charge "out there." All it knows is its own charge. The source then sends out a "signal" that radiates away from it radially, and this signal carries with it the information of how much charge the source has and how far the signal travels – the signal gets weaker as it gets farther from the source, because it spreads out on the surface of an ever-growing sphere. Now if another charge happens to be in the space near where this source is, it "receives" the signal, and it takes from it the information about the amount of charge of the source, as well as the signal strength itself (which includes the inverse-square-law separation information), and the direction from which the signal is coming. The affected charge puts all this information together with its own charge to determine the electric force it feels.

This "signal" is constantly emitted, so it is always everywhere in the space around the source charge, and it is called the *electric field* of that source charge. Since the signal carries information about both a magnitude (source charge and distance) and a direction (coming from the source charge's position), it essentially associates a vector with every point in space. Defining (as usual) the origin to be at the position of the source charge, the electric field vector at a specific point (defined by the position vector $\overrightarrow{r} = r\,\hat{r}$) due to the source charge is:

$$\overrightarrow{E} = \frac{k\,q_{source}}{r^2}\,\hat{r} \tag{1.2.1}$$

To visualize what the complete field looks like, imagine all of space filled with vectors. For a positive point charge, the vectors all point directly away from it, and the magnitudes of the vectors drop-off in length as they get farther from the source:

Figure 1.1.2 – Electric Field of a Point Charge



Comparing our Coulomb field equation with Equation 1.1.3, we see that indeed all of information needed to compute the electric force, except for the amount of charge that is affected, is contained within the electric field vector. So if we know the electric field vectors everywhere in space (or, more succinctly, we "know the electric field"), then we can compute the force on a point charge placed at any position, simply by multiplying the affected charge by the electric field vector:

$$\overrightarrow{F}_{on\ q} = q\,\overrightarrow{E}\,(r)\,, \qquad \text{where } \overrightarrow{r} = \text{position vector of the charge } q \tag{1.2.2}$$

Alert

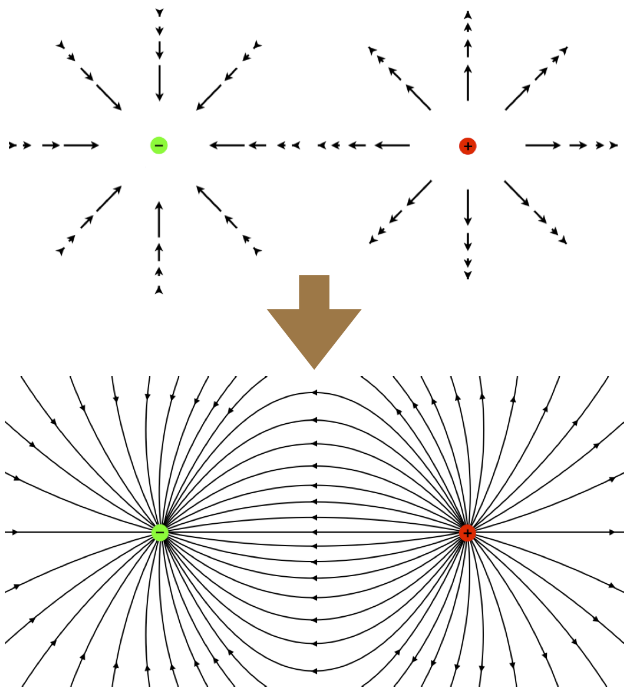
It is a common mistake to think that the electric field vector points in the direction of the force acting on a charge, but the affected charge can be either positive or negative. If it is a positive charge, then the direction of the force on it will be the same as the direction of the field, but if the affected charge is negative, then the force and field will point in opposite directions.

Field Superposition

While the field model may be only slightly more satisfying than the direct action model from a philosophical standpoint, it actually has some very pragmatic uses. The foremost of these is that it allows us to talk about forces on a particle without having to actually worry about the specifics of all the other particles affecting it. Once we determine the electric field due to one or a collection of charges, we can forget about those charges and just work with the field. Wait… “collection of charges?” How do we determine the electric field of a collection of charges? It turns out that we can superpose field vectors to get a single, aggregate field vector. That is, we don’t have to worry about how a charge is affected by the electric fields of a whole lot of other charges – we can instead aggregate the electric fields of all those charges, and call it *the* electric field in that space, and use it alone to determine the force on the charge in question. To compute the electric field contributions of several charges at a single point in space, we naturally have to add them *like vectors*.

While it is often useful to picture electric fields as collections of lots of little vectors of varying length and direction filling all of space, there is also another descriptive way to think about electric fields. This is called *electric field lines of force*, or simply, *electric field lines*. What this picture does is to merge the field vectors together, so that lines are created that point the direction of the field everywhere in space. So for example, consider two point charges with equal magnitudes and opposite signs in the same region of space (such a configuration is called a *dipole*, while a single point charge is called a *monopole*):

Figure 1.2.2 – Dipole Field Lines



The field line description clearly displays the *direction* of the electric field vectors everywhere in space (tangent to the lines drawn, in the direction of the arrows). One might suppose that it is nonetheless inferior to the many-little-arrows picture, in that it doesn't display the s*trength* of the field everywhere, but this is not correct. We know that the electric field get stronger as they get closer to the point charges, and there is a property of electric field lines that represents this – the *density* of field lines. The closer together the lines are, the stronger the field.

Consider the electric field lines for the dipole shown above along the horizontal axis joining the two charges. To the right of the positive charge, we notice that adjacent field lines are diverging from each other *faster* than if the positive charge was by itself (when they would emanate directly radially outward). With the density of the field lines indicating field strength, this means that the field is getting weaker in that direction *faster* than if the positive charge was the only source of the field. We already know how fast the field weakens with distance for a monopole (given in Equation 1.2.1), and in Section 1.4 we will approximate (for large values of r compared to the dipole charge separation) the faster rate at which this dipole field weakens with distance.

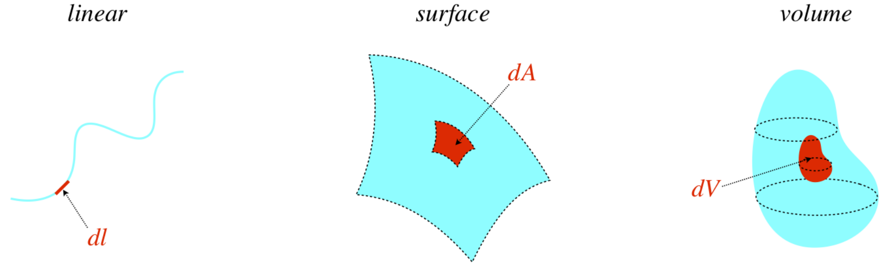
1.3: Computing Electric Fields for Known Charge Distributions

Continuous Charge Distributions

Charges occur in nature in discrete microscopic pieces. We can imagine a long line of charges, and the field that the superposition of fields of all these charges would create. If the charges are small enough and close enough together, the line of charge would look continuous. There are many examples of continuous collections of charge besides a linear string. For example, the string of charges doesn't need to be in a straight line. Also, it doesn't need to be distributed on a line at all, but can instead be spread across a surface or inside of a volume. These lines, surfaces, and volumes can take on an infinite number of shapes, although mathematically we can only really solve for the electric fields of the simpler, more symmetric ones, like straight lines, circles, planes, cylinders, and spheres.

The one aspect to solving such problems common to all of them is that in every case, we cease talking about individual charges and instead work with *charge density*. We discussed the idea of density in detail in both Physics 9A (e.g. mass density is used in calculating center of mass and rotational inertia) and Physics 9B (e.g. mass density of a medium affects the speed of a wave through it, and the density of a fluid plays a role in buoyancy). Here we are measuring charge-per-unit-whatever (the "whatever" being determined by the dimension into which the charge is distributed).

Figure 1.3.1 – Linear, Surface, and Volume Densities



The small (infinitesimal) bit of charge residing in each of these regions is found by multiplying the density (located at the position of that region – the density doesn't have to be the same everywhere!) by the size of the region. Typically linear charge densities are represented by a λ, surface charge densities by a σ, and volume charge densities by a ρ:

$$dq = \lambda \, dl \qquad dq = \sigma \, dA \qquad dq = \rho \, dV \tag{1.3.1}$$

Alert

The charge dq has units of coulombs in every case, which means that these three types of density all have different units: Cm^{-1}, Cm^{-2}, *and* Cm^{-3}*, respectively.*

If we want to know the total amount of charge along a line, on a surface, or within a volume, then we need to integrate these infinitesimal amounts over the full region. Again, the densities are not necessarily uniform, so the density is a function of position, and cannot be removed from the integral:

$$Q_{tot} = \int\limits_{whole\ line} \lambda \, dl \qquad Q_{tot} = \int\limits_{whole\ surface} \sigma \, dA \qquad Q_{tot} = \int\limits_{whole\ volume} \rho \, dV \tag{1.3.2}$$

When it comes to using continuous charge distributions to compute electric fields, there is one critical idea that comes into play:

An infinitesimal chunk of charge is so small that its field is precisely that of a point charge, so we can write it as a coulomb field.

So the idea of computing the electric field of a continuous distribution of charge is to write down the (coulomb) electric field of a single infinitesimal chunk, and then add up (integrate) the electric field contributions of all the chunks.

Alert

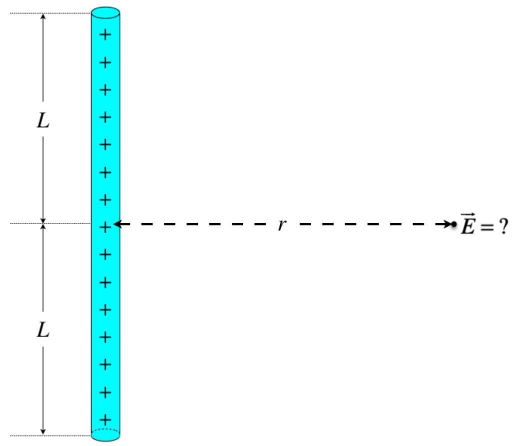
We have to keep in mind that the electric field contributions are vectors, which adds a layer of complication to the problem. The right approach is to compute the components of the field vector separately. Fortunately, there are often tricks related to symmetry we can use to simplify this particular step.

The secret to solving such problems is to proceed slowly and methodically. We will break down the steps here with a simple example of a 1-dimensional charge distribution.

Electric Field of a Straight Line Segment of Uniform Density

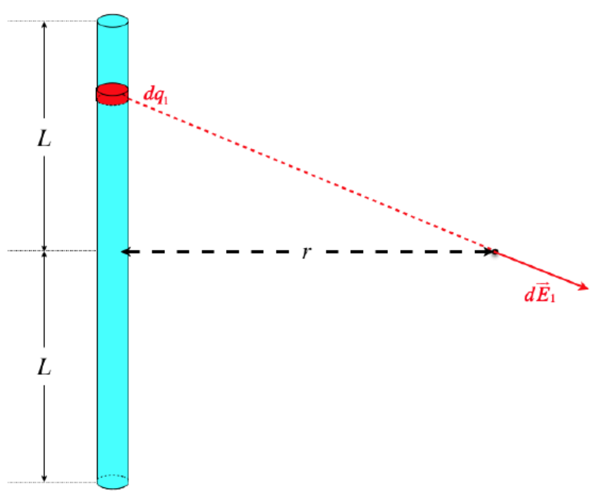
Our line segment will have a length of $2L$, a total charge of Q, and we will be computing the field at a position a perpendicular distance r from the center of the line segment. We'll treat the charge as if it is positive. If we want the solution for a negative charge, we just have to flip the direction of the electric field.

Figure 1.3.2a – Field of a Uniform Line Segment



Step 1: Sketch an arbitrary infinitesimal chunk of charge, and its contribution to the field vector.

Figure 1.3.2b – Field of a Uniform Line Segment



Step 2: Break the field vector into appropriate components and make a symmetry argument (if applicable).

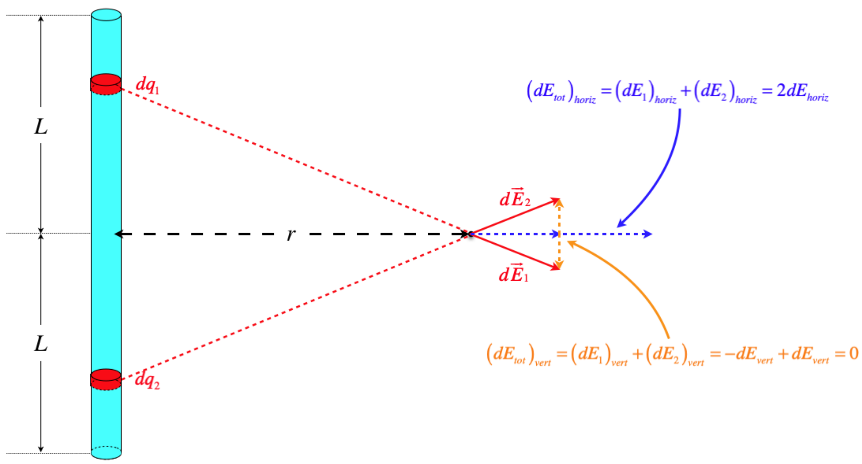
This step requires a bit of clarification. Our integral will be adding field contributions for every little chunk of charge in the line segment, but we have to add these as vectors. The simplest way to do this is to add like components. In this case, the electric field of every chunk will generally have a vertical and horizontal component (within the plane of the page). We will therefore need to do *two* integrals – one that adds up the vertical electric field components, and one that adds up the horizontal components.

Notice that for every chunk above the center line, there is another corresponding chunk below it. These "twin" chunks will provide horizontal components to the field that are equal and in the same direction, but the vertical components, while equal in magnitude, *oppose each other*. When we add up the vertical components over the entire line segment, we therefore find that they cancel, meaning we know (without performing an integral) that the electric field has zero vertical component.

Alert

This symmetry only applies to points on the line perpendicular to the center of the line segment, and only because the charge distribution is uniform. Fortunately, that is what we are calculating here.

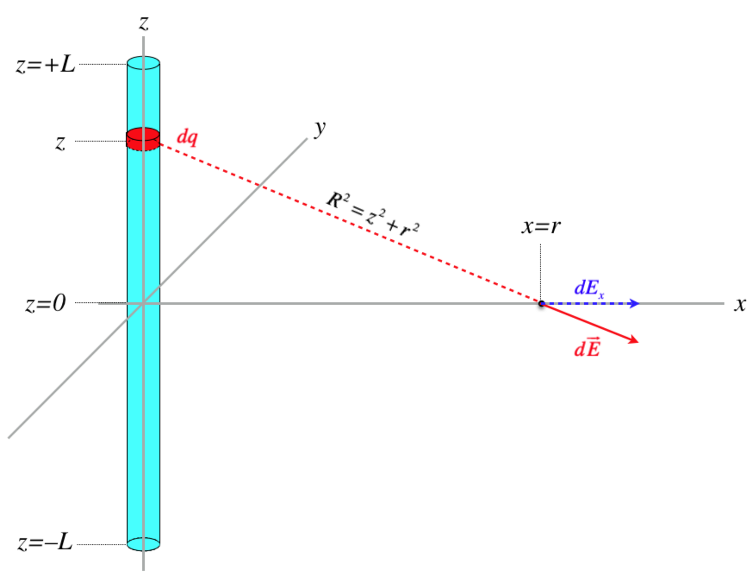
Figure 1.3.2c – Field of a Uniform Line Segment



Step 3: Introduce a coordinate system and label everything.

Don't be afraid to introduce variables! Ultimately the answer needs to be in terms of what has been given (in this case, that would be L, Q, and r), but since you can't simply write down the answer, you need to create some variables to work with. Besides the coordinate axes, we introduce the variable R to represent the distance separating a chunk of charge from the position of the field, and the angle θ to allow us to express the horizontal component of the field. Note that the $y-$axis in the diagram below is not necessary, as it is clear that the field will not have a $y-$component in the coordinate system shown.

Figure 1.3.2d – Field of a Uniform Line Segment



Step 4: Relate the differential chunk of charge to the charge density, using the coordinate system.

This is a linear distribution and the length of the chunk expressed in terms of the coordinate system is dz, so we have:

$$dq = \lambda \, dz \tag{1.3.3}$$

But we also know that the charge density is uniform, so it is the ratio of the total charge and the total length:

$$\lambda = \frac{Q}{2L} \quad \Rightarrow \quad dq = \frac{Q}{2L} \, dz \tag{1.3.4}$$

Step 5: Apply the fact that an infinitesimal chunk acts like a point charge to produce an infinitesimal coulomb field.

We have called the distance from the chunk to the point of the field R, so:

$$dE = \frac{k \, dq}{R^2} = \frac{kQ}{2L}\left[\frac{dz}{z^2+r^2}\right] \tag{1.3.5}$$

Step 6: Construct the component of the field from the magnitude, and convert variables into the integration variable.

We can't just integrate dE at this point! We are adding the horizontal components, not the magnitudes. With our definition of θ, we have:

$$dE_x = dE\cos\theta = \frac{kQ}{2L}\left[\frac{dz}{z^2+r^2}\right]\cos\theta \tag{1.3.6}$$

We still can't integrate yet, because as we start accounting for different chunks (i.e. integrate over z), the value of θ changes. We therefore need to write θ in terms of z. Fortunately, we have a nice right triangle to work with:

$$\cos\theta = \frac{r}{R} = \frac{r}{\sqrt{z^2+r^2}} \quad \Rightarrow \quad dE_x = \frac{kQr}{2L}\left[\frac{dz}{\left(z^2+r^2\right)^{\frac{3}{2}}}\right] \tag{1.3.7}$$

Step 7: Integrate over all of the chunks of electric charge.

In this case, the chunks lie from $z=-L$ to $z=+L$. We already noted that symmetry demands that the contributions of the lower half will be the the same as those of the upper half. We can see that this will be the case by replacing z in the integrand with $-z$ – the same integral results. If we like, we can change the limits to $0 \rightarrow +L$ and multiply the integral by 2:

$$E_x = 2\int_0^L \frac{kQr}{2L}\left[\frac{dz}{\left(z^2+r^2\right)^{\frac{3}{2}}}\right] \tag{1.3.8}$$

Perhaps your palms are getting sweaty at this point, with the thought of now having to perform an integral like this. Well, first of all, it isn't that bad. But even if it was, we will be more focused on the *physics* of these problems than proficiency with techniques of integration, so we will liberally use integral tables to look these up once we have constructed them. This is not to say that you never have to do any math, of course. It's not always obvious how the integral you have can be placed in the form given in an integral table, so you are expected to make substitutions or do whatever it takes to make the integral tables actually *useful* to you. Anyway, this particular integral comes out to be (check it by taking a derivative!):

$$E_x = \frac{kQr}{L}\left[\frac{z}{r^2\sqrt{z^2+r^2}}\right]_0^L = \frac{kQ}{r\sqrt{L^2+r^2}} \quad \Rightarrow \quad \overrightarrow{E}(r) = \frac{kQ}{r\sqrt{L^2+r^2}}\hat{r} \tag{1.3.9}$$

Notice that after the integral is performed, all of the variables we introduced are gone (the last one to vanish being the integration variable).

We already concluded that the y and z components of the field vanish at this point. The y component is zero because we defined the coordinate system so that this would be true. If we were to treat this more generally, we would note that no matter how we set up the x and y axes, the electric field points directly away from the segment (axially outward for a line of positive charge), which is why we have used the unit vector $\hat{r}$ instead of the unit vector along the x-axis only.

The z component is zero because of symmetry, but if we hadn't noticed this shortcut, we could still have concluded this. Going back to step 5, we would compute:

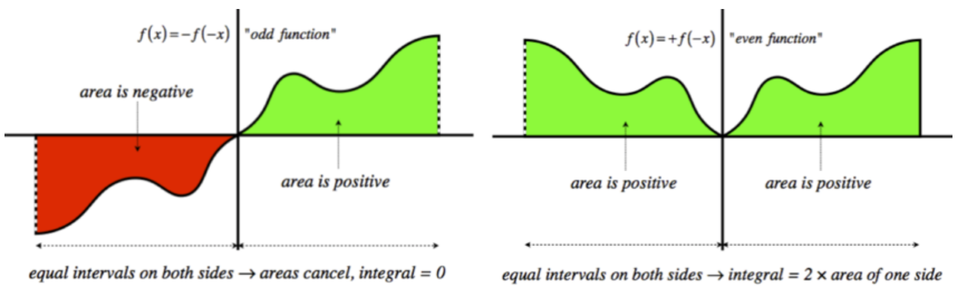
$$dE_z = dE\sin\theta = \frac{kQ}{2L}\left[\frac{dz}{z^2+r^2}\right]\sin\theta \tag{1.3.10}$$

Using our right triangle to substitute for $\sin\theta$, we get a z in the numerator instead of the r that we got last time:

$$dE_z = \frac{kQ}{2L}\left[\frac{zdz}{(z^2+r^2)^{\frac{3}{2}}}\right] \tag{1.3.11}$$

This integral differs from the previous one by a factor of z, which makes all the difference. Once again, the limits of integration are symmetric about the origin, but now if we replace z in the integrand with $-z$, we get the *negative* of the original integral. When these are added together for the two halves of the integral, we get zero. This is a nice mathematical trick you can use in these situations that avoids the need to do integrals like this one that come out to zero. If the integral is being performed over an interval over which the integrand is "odd" about the center of that interval (I will define "odd/even" in this context in a moment), then the integral comes out to be zero. If the integrand is "even," then the integral is double the value of the integral over half the interval. The integrand above is an odd function of z, meaning the function changes sign when $z \to -z$. The limits of integration are symmetric across the origin, so this fits the bill for a zero integral as described.

Figure 1.3.3 – Integrating Odd vs. Even Functions

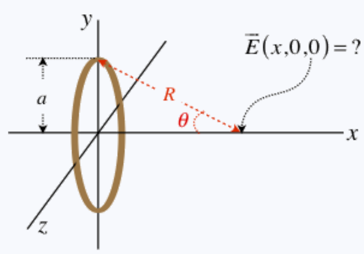


Example 1.3.1

A thin circular plastic ring with a radius of a carries a net charge Q that is uniformly distributed throughout. For a common point of reference, we will place this ring in the yz plane, centered at the origin. Find the electric field at all points on the x-axis.

Solution

First a diagram, to label the needed variables:



The distance from a small segment of the circle to the point x is:

$$R = \sqrt{a^2+x^2}$$

So the small contribution to the magnitude of the electric field by this element is:

$$dE = \frac{k\,dq}{a^2+x^2}$$

Thanks to the symmetry of the circle, the y and z components of this electric field all cancel out (for every element on the ring there is an element on the opposite side that cancels the components that are not in the $x-$direction). So the only part of the electric field we can keep is the $x-$component, which we find by multiplying the magnitude by the cosine of the angle θ shown in the diagram. This cosine can be written in terms of the right triangle shown above:

$$dE_x = dE\cos\theta = \left(\frac{k\ dq}{a^2+x^2}\right)\left(\frac{x}{R}\right) = \frac{kx\ dq}{\left(a^2+x^2\right)^{\frac{3}{2}}}$$

Now all we have to do is add up all of the dq contributions. The amount of charge in an infinitesimal segment of the circle (which is given by the radius times the infinitesimal angle the segment subtends) in terms of the linear charge density λ is:

$$dq = \lambda\left(a\ d\phi\right)$$

Because the charge density is uniform, it is simply equal to the total charge divided by the length, which is the circumference of the loop:

$$\lambda = \frac{Q}{2\pi a} \quad \Rightarrow \quad dq = \frac{Q}{2\pi}d\phi$$

Plugging this in and integrating around the entire ring (ϕ ranges from 0 to 2π, and x remains constant throughout the integral) gives the answer:

$$E_x = \frac{Q}{2\pi}\frac{kx}{\left(a^2+x^2\right)^{\frac{3}{2}}}\int_0^{2\pi} d\phi = \frac{kQx}{\left(a^2+x^2\right)^{\frac{3}{2}}}$$

This is just the $x-$component of the field, but since the other components vanish, we have that the field vector on the axis at point x is:

$$\overrightarrow{E}\left(x,0,0\right) = \frac{kQx}{\left(a^2+x^2\right)^{\frac{3}{2}}}\hat{i}$$

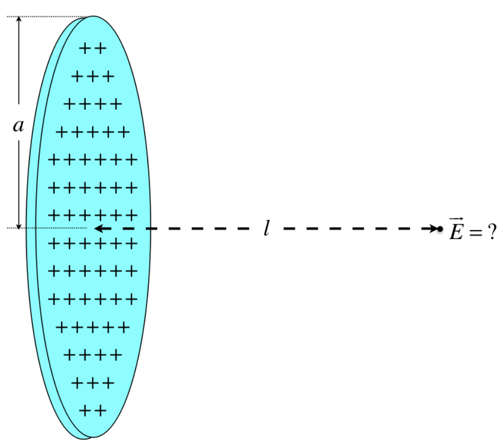
Note that the direction changes when x is negative, as it should. Also note that the electric field vanishes at the origin and at infinity, as we would expect.

Electric Field of a Circular Disk of Uniform Density

We next take on a problem that differs from the previous one in two ways. First, it involves a two-dimensional charge distribution (which requires a double integral), and second, we will need to use polar coordinates rather than cartesian. Despite the differences, however, the steps we take are the same.

Our disk will have a radius of a, a total charge of Q, and we will be computing the field at a position a perpendicular distance l from the center of the disk.

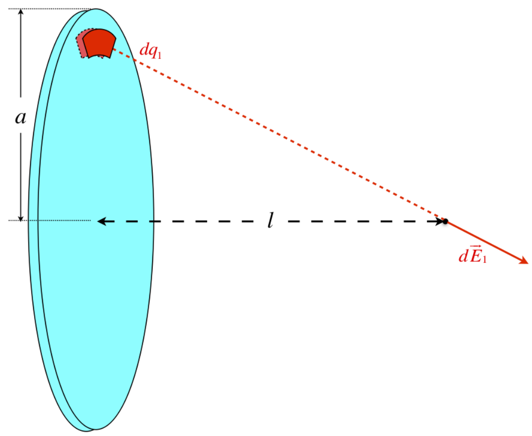
Figure 1.3.4a – Field of a Uniform Disk



Step 1: Sketch an arbitrary infinitesimal chunk of charge, and its contribution to the field vector.

This is a bit trickier than before, because the chunk is not a simple line segment, and we have to keep an eye on what we will be integrating later. For example, we could draw a chunk that is a square, but integrating this in cartesian coordinates would be challenging. So instead, we choose an arc with a radial thickness.

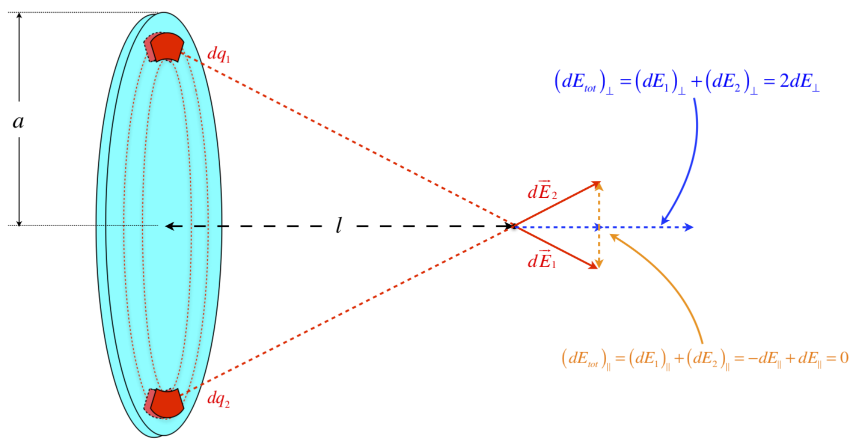
Figure 1.3.4b – Field of a Uniform Disk



Step 2: Break the field vector into appropriate components and make a symmetry argument (if applicable).

As before, we find that there is symmetry (this time, axial) which creates opposing contributions to the electric field vector along directions perpendicular to the central axis.

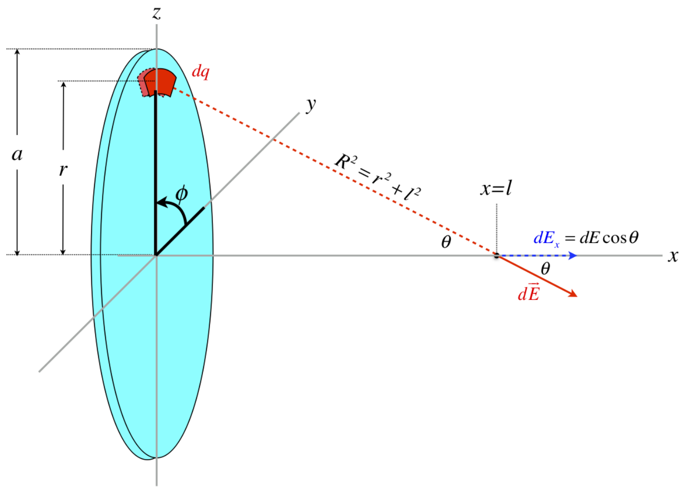
Figure 1.3.4c – Field of a Uniform Disk



Step 3: Introduce a coordinate system and label everything.

The shape of our collection of charge calls for a use of polar coordinates to describe the position of the chunks of charge.

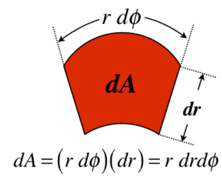
Figure 1.3.4d – Field of a Uniform Disk



Step 4: Relate the differential chunk of charge to the charge density, using the coordinate system.

This is a surface charge density, so we need to multiply the constant charge density by the infinitesimal area of the chunk in order to get dq. Since we are using polar coordinates, this requires a closer look at the geometry of the chunk:

Figure 1.3.4e – Field of a Uniform Disk



The infinitesimal charge in terms of the charge density is therefore:

$$dq = \sigma\ dA = \sigma r dr d\phi \tag{1.3.12}$$

The charge density is uniform, which means that the total charge is the density multiplied by the total area of the disk, which gives:

$$\sigma = \frac{Q}{\pi a^2} \quad \Rightarrow \quad dq = \frac{Q}{\pi a^2} r\ dr\ d\phi \tag{1.3.13}$$

Step 5: Apply the fact that an infinitesimal chunk acts like a point charge to produce an infinitesimal coulomb field.

$$dE = \frac{k\ dq}{R^2} = \frac{kQ}{\pi a^2}\left[\frac{r\ dr\ d\phi}{r^2+l^2}\right] \tag{1.3.14}$$

Step 6: Construct the component of the field from the magnitude, and convert variables into the integration variable.

$$dE_x = dE\cos\theta = \frac{kQ}{\pi a^2}\left[\frac{r\ dr\ d\phi}{r^2+l^2}\right]\cos\theta = \frac{kQ}{\pi a^2}\left[\frac{r\ dr\ d\phi}{r^2+l^2}\right]\left[\frac{l}{R}\right] = \frac{kQl}{\pi a^2}\frac{r\ dr\ d\phi}{(r^2+l^2)^{\frac{3}{2}}} \tag{1.3.15}$$

Step 7: Integrate over all of the chunks of electric charge.

To include the contributions of all the chunks, we integrate the angle around the entire circle ($\phi = 0 \to 2\pi$), and the radius from the center out to the edge ($r = 0 \to a$). Notice that integrating around the whole circle and from the center to the radius takes care of the "opposing" chunks everywhere around the disk – the components of the field perpendicular to the axis cancel, and those parallel all add (no need to multiply by 2, as we did when we integrated from the center of the line of charge earlier, to include both sides).

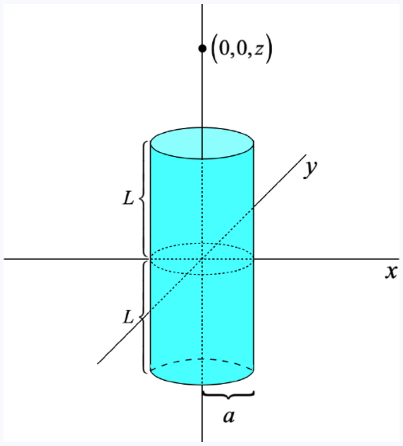
$$E_x = \frac{kQl}{\pi a^2}\int_0^{2\pi} d\phi \int_0^a \frac{r\ dr}{(r^2+l^2)^{\frac{3}{2}}} \tag{1.3.16}$$

The integral over $d\phi$ is simply equal to 2π, and the integral over r is so common that we shouldn't even have to look it up (though it is fine to do so). Make the substitution $u \equiv \left(r^2+l^2\right)^{-\frac{1}{2}}$ and the integral comes out immediately:

$$E_x = \frac{kQl}{\pi a^2}(2\pi)\left[\frac{-1}{\sqrt{r^2+l^2}}\right]_{r=0}^{r=a} \quad \Rightarrow \quad \overrightarrow{E}(l,0,0) = \frac{2kQ}{a^2}\left(1-\frac{l}{\sqrt{a^2+l^2}}\right)\hat{i} \tag{1.3.17}$$

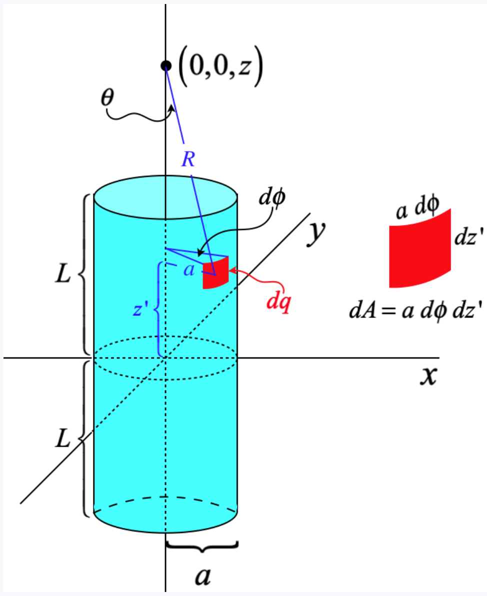
Example 1.3.2

A thin, hollow, plastic cylinder with a radius of a and length $2L$ carries a net charge Q that is uniformly distributed throughout. For a common point of reference, we will place this cylinder with its axis along the z-axis, centered at the origin (see the diagram). Find the electric field at all points on the z-axis.



Solution

Every infinitesimal piece of this cylinder behaves like a point charge, and we need to add the contributions of those point charges to the electric field. So we start by diagramming an arbitrary element of charge as a small patch of area on the cylinder:



The infinitesimal patch of charge has an area that equals its height dz' multiplied by its arclength $ad\phi$, as indicated in the diagram.

[Note that the z-position of the patch (and its infinitesimal length in the z-direction) are labeled with a prime (z'). This is because we will be adding up the contributions of all the patches, which entails integrating over that variable, and we

need to distinguish that quantity from the value z that describes the position where we are measuring the electric field.]

To calculate the electric field requires calculating all three components separately, but we find once again that two of the components (E_x and E_y vanish thanks to the symmetry we have by considering only points on the z-axis. We therefore only calculate the z component contributed by every charge element, which, in terms of the electric field magnitude of the charge element is:

$$dE_z = dE\cos\theta = dE\left(\frac{side\ adjacent}{hypothenuse}\right) = dE\left(\frac{z-z'}{R}\right)$$

Now we write down the coulomb field for the point charge, once again writing the charge element in terms of the constant surface density, which we will call σ:

$$dE = \frac{kdq}{R^2} = \frac{k\sigma dA}{R^2} = \frac{k\sigma a\ d\phi dz'}{R^2}$$

The charge density is uniform, so it is simply the total charge divided by the total surface area of the cylinder:

$$\sigma = \frac{Q}{2\pi a\,(2L)} = \frac{Q}{4\pi aL} \tag{1.3.18}$$

Putting together everything we have so far gives:

$$dE_z = \frac{kQ\,(z-z')}{4\pi LR^3}\,d\phi dz'$$

We are going to have to integrate over the entire surface of the cylinder, so we need to integrate over the variables z' and ϕ. No part of the integrand depends upon ϕ, and the limits of integration for that variable are $0 \to 2\pi$. The quantity R depends upon z', which we can express using the pythagorean theorem. The limits for the integration over z' are $-L \to +L$, so we are left with:

$$E_z = \frac{kQ}{4\pi L}\int_0^{2\pi} d\phi \int_{-L}^{+L} \frac{z-z'}{\left[a^2+(z-z')^2\right]^{\frac{3}{2}}}dz'$$

The integral over ϕ is trivial, and the second integral is pretty straightforward with the substitution $u \equiv z-z'$:

$$E_z = \frac{kQ}{2L}\int_{z+L}^{z-L} \frac{-u}{\left[a^2+u^2\right]^{\frac{3}{2}}}du = \frac{kQ}{2L}\left[\frac{1}{\sqrt{a^2+u^2}}\right]_{z+L}^{z-L} = \frac{kQ}{2L}\left[\frac{1}{\sqrt{a^2+(z-L)^2}} - \frac{1}{\sqrt{a^2+(z+L)^2}}\right]$$

Example 1.3.3

Show that the solutions to the two previous examples are consistent with each other, by noting that a thin ring is equivalent to a very hollow cylinder with a very short length.

Solution

We could get the solution to Example 1.3.1 by taking the limit of the solution to Example 1.3.2 as the length of the cylinder goes to zero:

$$\lim_{L\to 0} E_z = \frac{kQ}{2\cdot 0}\left[\frac{1}{\sqrt{a^2+z^2}} - \frac{1}{\sqrt{a^2+z^2}}\right] = \frac{0}{0}$$

We get an indeterminate form, which means we need to use l'Hôpital's rule:

$$\lim_{L\to 0} E_z = \frac{kQ}{2}\lim_{L\to 0}\frac{\frac{d}{dL}\left[\frac{1}{\sqrt{a^2+(z-L)^2}}-\frac{1}{\sqrt{a^2+(z+L)^2}}\right]}{\frac{d}{dL}L}$$
$$= \frac{kQ}{2}\lim_{L\to 0}\left[\frac{z-L}{\left[a^2+(z-L)^2\right]^{\frac{3}{2}}}+\frac{z+L}{\left[a^2+(z+L)^2\right]^{\frac{3}{2}}}\right]$$
$$= \frac{kQz}{(a^2+z^2)^{\frac{3}{2}}}$$

This is the same result as in Example 1.3.1, where, of course, we have swapped the x-axis for the z-axis.

Infinite Lines and Planes of Charge

A particularly useful aspect of the solutions provided above for the line segment and the disk lies in extending the result to charge distributions that are infinite in extent. To see how to make this extension, we first need to change the form of the solutions so that they reflect the charge density, rather than the total charge:

$$\overrightarrow{E}_{line\ segment}(r) = \frac{kQ}{r\sqrt{L^2+r^2}}\hat{r} = \frac{2kL\lambda}{r\sqrt{L^2+r^2}}\hat{r} \tag{1.3.19}$$

$$\overrightarrow{E}_{disk}(x) = \frac{2kQ}{a^2}\left(1-\frac{x}{\sqrt{a^2+x^2}}\right)\hat{i} = 2\pi k\sigma\left(1-\frac{x}{\sqrt{a^2+x^2}}\right)\hat{i} \tag{1.3.20}$$

For both cases, we can extend the solution to the infinite cases by letting the parameter a go to infinity.

$$\overrightarrow{E}_{line}(r) = \lim_{L\to\infty}\frac{2kL\lambda}{r\sqrt{L^2+r^2}}\hat{r} = \lim_{L\to\infty}\frac{2k\lambda}{r\sqrt{1+\frac{r^2}{L^2}}}\hat{r} = \frac{2k\lambda}{r}\hat{r} \tag{1.3.21}$$

$$\overrightarrow{E}_{plane}(x) = \lim_{a\to\infty} 2\pi k\sigma\left(1-\frac{x}{\sqrt{a^2+x^2}}\right)\hat{i} = 2\pi k\sigma\hat{i} \tag{1.3.22}$$

The solutions for the infinite line and plane take on significantly simpler forms than for the finite cases. But there is another simplification for these cases as well: In the finite cases, we had to be careful to remember that the solutions only applied along the point of symmetry, equally-spaced between the extremes of the charge distributions (i.e. in the xy plane for the line segment, and on the $x-$axis for the disk). But with these distributions now having infinite extent, *everywhere* represents a symmetric position, so these solutions are good everywhere in space.

It might seem strange that the electric field is *uniform everywhere in space* for the infinite plane. Why doesn't the field get stronger closer to the plane of charge? This can be more easily seen using the field line description. If the field got stronger closer to the plane, then the field lines would have to get closer together there. But that can only happen if the field lines are not perpendicular to the plane everywhere. Due to the symmetry of the infinite plane, there is no reason to believe that the field would have any y or z components anywhere in space. With the field only pointing along the x direction, the field lines can't change their separation, and the field strength remains constant everywhere.

One might ask that although these solutions have simpler forms, how can they be "useful," as claimed above? How many infinite lines or planes of charge does one run across in real life? The answer is that *all* solutions in physics are approximations, and are only useful up to the sensitivity of the measurements. For example, if we look at the field of a finite-sized plane of charge, but look at a position in space that is very close to the plane compared to the dimensions of that plane, then treating it as "infinite" is a good approximation. What is especially nice about this approximation is that so long as we are looking at positions close to the plane compared to the distance from the edges, we don't even care what shape the plane is – it can be a circular disk, a square, or some random, jagged shape.

One other comment about usefulness. Once we have made a computation, we don't have to do it over again. That means that to the extent that a new problem has elements that can be approximated as these above, we can simply use the solutions – we don't have to start from scratch and perform an integral every time. We saw this utility once before in Physics 9A, where we may have performed an integral to compute the moment of inertia of a rod or some other object, but after that, we can just use what we computed.

Non-Uniform Charge Distributions

So far all we have considered are charge distributions that are uniform. These cases involve two simplifications. The first is that the physical situation is more likely to include a symmetry that simplifies the calculation of the field. For example, if the case of the line segment of charge solved above did not include a uniform charge density, then the electric field on the x-axis may not have a vanishing z-component. Second, the relationship between the charge and charge density is different, which puts another function into the integral. Despite these differences, there are still cases for which symmetry is still a useful tool, and the overall approach is exactly the same as has been shown above.

Example 1.3.4

A positively-charged thin plastic rod of length L is placed on the $+x$-axis with one end at the origin. The linear distribution of charge on this rod satisfies the equation:

$$\lambda\left(x\right)=\lambda_o\frac{x^2}{L^2}$$

A point particle with the same charge as the rod is also placed somewhere on the $-x$-axis, and the resulting total electric field at the origin vanishes.

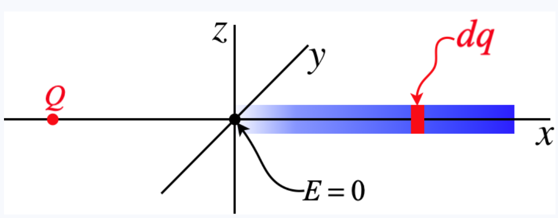
a. Find the charge of the point particle (in terms of the given quantities L and λ_o).
b. Find the position of the point particle.

Solution

a. The point particle has the same total charge as the rod, so we need to compute the charge on the rod given the density function. Unlike cases of uniform linear charge distributions where the total charge is simply the product of the density and the length, in this case, we need to add up all the small contributions:

$$Q=\int dq=\int\lambda\left(x\right)dx=\int_0^L\lambda_o\frac{x^2}{L^2}dx=\lambda_o\left[\frac{x^3}{3L^2}\right]_0^L=\frac{1}{3}\lambda_oL$$

b. Now we need to compute the electric field due to the rod in terms of the charge, and then we can determine the position of the point particle that will cancel that field. We start with a diagram:



The amount of charge contained in the infinitesimal slice is, as we used above:

$$dq=\lambda dx=\lambda_o\frac{x^2}{L^2}dx$$

There is no need to worry about components of the electric field here, as it clearly only has an x-component at the origin. Calling the position of the tiny bit of charge x, the field at the origin due to dq is the usual coulomb field, giving:

$$dE=\frac{kdq}{x^2}=\frac{k\left[\lambda_o\frac{x^2}{L^2}dx\right]}{x^2}\quad\Rightarrow\quad E=\frac{k\lambda_o}{L^2}\int_0^L dx=\frac{k\lambda_o}{L}$$

We need this field in terms of the total charge, so we can use the result of part (a) to substitute for λ_o:

$$\lambda_o = \frac{3Q}{L} \quad \Rightarrow \quad E = \frac{3kQ}{L^2}$$

The electric field of the point particle must have this same magnitude at the origin, so calling its distance from the origin r we find:

$$\frac{kQ}{r^2} = \frac{3kQ}{L^2} \quad \Rightarrow \quad r = \frac{L}{\sqrt{3}}$$

Some General Comments

Now for a few general observations about this process of computing the electric field…

1. The sub-steps become tougher as the geometry becomes more complex. Here are a few of these problem spots:

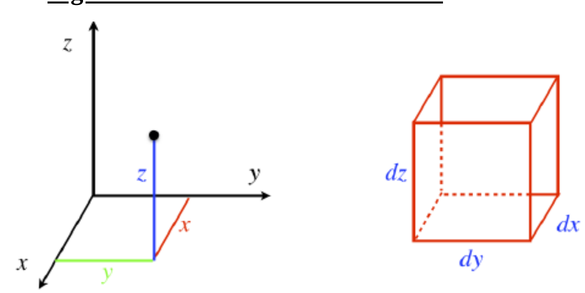
- Labeling the position of the source point charge is a bit trickier if it lies within a 2-dimensional plane (or within a 3-dimensional volume), rather than on a 1-dimensional line. For example, for a surface of charge that is a circle or a cylinder, you *could* use rectangular coordinates, but then the limits of integration get significantly tougher. In general, the symmetry of the physical situation will dictate what is the best coordinate system to use.
- The element of charge written in terms of the coordinates is more complicated. Typically we just choose an appropriate coordinate system and look up the expression for an infinitesimal element of length/area/volume (see below for references), but that doesn't mean it's easy.
- For surfaces, there are two variables over which we must integrate (for volumes, where we will not venture, there are three). If we choose a good coordinate system, symmetry usually means that we have only one non-trivial integral. Also, we could choose an element that allows for a shortcut (such as that which we used in Example 1.3.3), but that gets away from the notion of superposing the fields of many point dq charges, which can confuse our understanding of the situation on physical grounds.

1. We have relied heavily upon symmetry to avoid doing calculations of many of the components that vanish, but one needs to be aware that in general solving for the electric field involves *three* solutions – one for each component – all performed in the same manner. It is a good exercise to go back to the examples above and show that the zero components are indeed zero without resorting to symmetry arguments. If the reader cannot manage this, then that is an indication of a shortcoming in understanding how such calculations are performed.
2. While every case shown above involves finding the field *outside* the charge distribution, there is also an electric field *within* the distribution, calculated exactly the same way.

Area and Volume Elements

For reference purposes, we present here the area and volume elements for cartesian, cylindrical, and spherical coordinates.

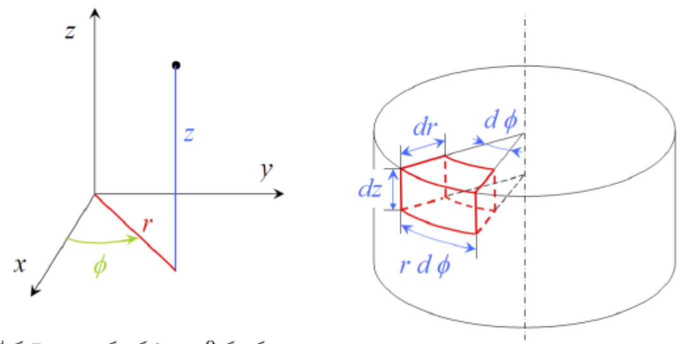
Figure 1.35 – Cartesian Coordinates



$$-\infty \le x \le +\infty, \quad -\infty \le y \le +\infty, \quad -\infty \le z \le +\infty$$

$$\textit{area element}: \quad dA = dx\,dy \;\; or \;\; dy\,dz \;\; or \;\; dz\,dx$$
$$\textit{volume element}: \quad dV = dx\,dy\,dz$$

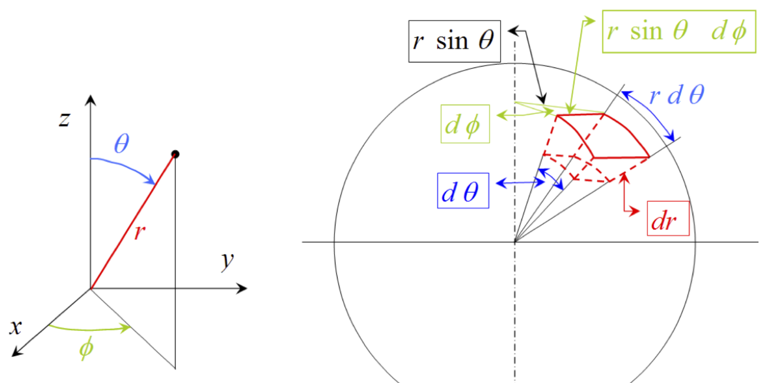
Figure 1.36 – Cylindrical Coordinates



$$0 \le \phi \le \pi, \quad -\infty \le z \le +\infty, \quad 0 \le r \le \infty$$

$$\begin{aligned}
&\textit{area element (top \& bottom)}: && dA = r\, d\phi\, dr \\
&\textit{area element (side)}: && dA = r\, d\phi\, dz \\
&\textit{volume element}: && dV = r\, d\phi\, dz\, dr
\end{aligned}$$

Figure 1.37 – Spherical Coordinates



$$0 \le \theta \le \pi, \quad 0 \le \phi \le 2\pi, \quad 0 \le r \le \infty$$

$$\begin{aligned}
&\textit{area element}: && dA = r^2 \sin\theta\, d\theta\, d\phi \\
&\textit{volume element}: && dV = r^2 \sin\theta\, d\theta\, d\phi\, dr
\end{aligned}$$

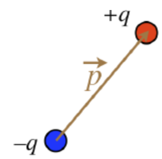
1.4: Dipoles

Definition of a Dipole

Whenever possible, it is a good idea to "package" common physical situations into models that we can re-use without having to always reinvent the wheel. We actually discussed this to some extent in the previous section, in the context of solving for the field of an infinite line or plane of charge. We know well how to deal with single point charges, but of course physical systems rarely behave in such a simple way. What we are going to look at here is a model for *two* point charges that are equal in magnitude and opposite in sign. The reason this is such an important package to develop is that it appears so much in nature, in the form of neutrally-charged molecules.

Consider two equal point charges, one positive, and the other negative, that are held rigidly at a fixed separation distance (if you like, you can imagine a tiny rigid rod holding them at fixed relative positions). We have already seen what the field of such a dipole configuration looks like, in Figure 1.2.2. We could forever treat such a configuration as a combination of two point particles, but it is helpful to package them so that we can treat them as a single entity and not have to go back and recalculate things. To that end, we define a vector quantity known as an *electric dipole moment* as follows:

Figure 1.4.1 – Electric Dipole Moment



The magnitude of the dipole moment is defined as the product of the absolute value of one of the two charges, multiplied by the distance separating the two charges:

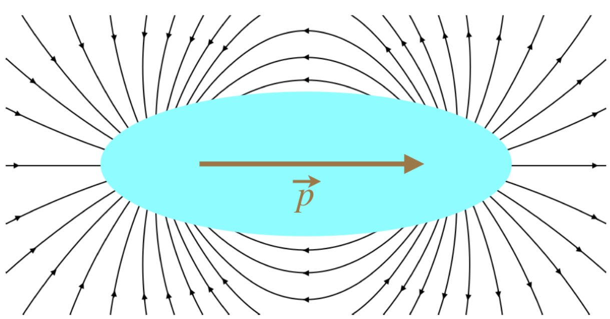
$$\left|\vec{p}\right| \equiv q\,d \tag{1.4.1}$$

The direction of the dipole moment is that it points *from the negative charge to the positive charge*.

> **Alert**
>
> *Chemists typically define the dipole moment as pointing in the opposite direction. When creating a "package" for later use, how you define it is up to you. We will see that there are compelling reasons (at least in physics applications) for defining it as above.*

Note that the dipole moment is not the same as the dipole electric field. It may seem funny to even mention this, as these two quantities are not even close to being the same, but it does come up. One place where it gets confusing is that the dipole moment points in the *opposite* direction as the electric field between the two charges. But as we are forming a package with these two charges, what happens *between* them is of no consequence. When it comes to the direction of the *dipole field*, the dipole moment direction makes perfect sense.

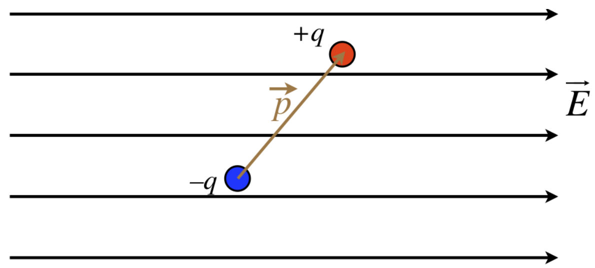
Figure 1.4.2 – Field of a Dipole



Dipoles in External Fields

We consider now the effect that a uniform electric field has on a dipole. Note that while we will be assuming a uniform field, in reality we mean that the amount that the external field changes across the length of the dipole is negligible. Also, as will generally be the case going forward, when we draw a diagram of a uniform field, we will represent it with a set of parallel field lines.

Figure 1.4.3 – Dipole in a Uniform Field

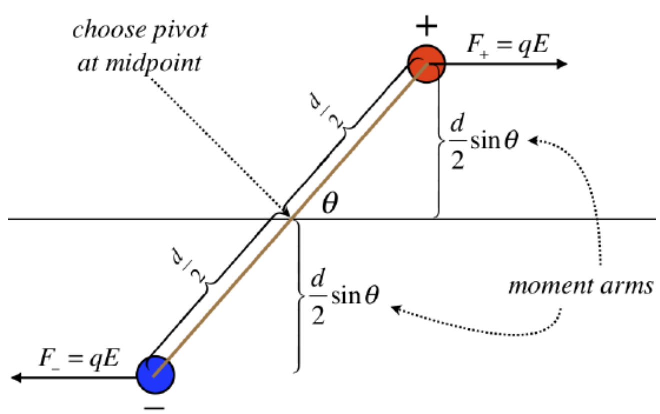


We begin by considering the force on the dipole. Certainly each individual charge feels a new force from the field, but the charges are equal in magnitude, and the forces act in opposite directions, so the net force on it is zero.

> **Alert**
>
> *If the field is not uniform, then the dipole can experience a net force! This might seem odd, given that the "package" of two charges is neutrally-charged, but it is an important physical effect to be aware of (we will discuss it in more detail later).*

With no net force, the center of mass of the dipole will not accelerate, but there will clearly be a *torque* exerted on this object. We can introduce a coordinate system above, and determine what this torque is, in terms of the field and dipole moment.

Figure 1.4.4 – Torque on a Dipole



Multiplying the forces by the moment arms, and summing, we find that the magnitude of the torque on this dipole is:

$$\tau = 2\left[qE\frac{d}{2}\sin\theta\right] = qd\;E\;\sin\theta \tag{1.4.2}$$

The magnitude of the dipole moment appears in the equation, as does the strength of the electric field, and the sine of the angle between them. This would suggest a connection to the cross product of the dipole moment and the electric field vector. Looking at the diagram, we see that the torque will cause clockwise rotational acceleration, which means that the torque vector points into the page. Indeed, the right-hand-rule applied to the cross product of $\overrightarrow{p}$ and $\overrightarrow{E}$ results in a vector that points into the page, so we conclude:

$$\overrightarrow{\tau} = \overrightarrow{p} \times \overrightarrow{E} \tag{1.4.3}$$

> **Example 1.4.1**
>
> *A dipole is a distance r from an infinitely-long line of negative charge of density λ.*

a. *The dipole moment $\vec{p}$ is parallel to the line of charge. Find the magnitude of the torque on the dipole in terms of r, $\vec{p}$, and λ.*
b. *The dipole moment is now pointing directly at the line of charge (perpendicular to it). Is there a net force on the dipole, and if so, is it toward or away from the line of charge?*

Solution

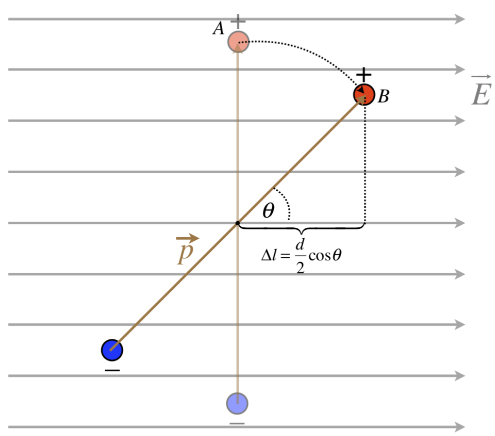
a. The field is the same strength at both ends of the dipole, so we can just use the torque equation. The field is axially inward to the line of charge, which means that it is perpendicular to the dipole moment, so plugging in the field of the long line of charge, we get:

$$\tau = p\ E\ \sin 90^o = \frac{2kp\lambda}{r}$$

b. The field from the long line of charge is not uniform – it is stronger closer to the line. Therefore the "front" of the dipole, which is closer to the line of charge, feels a strong force than the "rear." The line of charge is negative, which means the front of the dipole (which is the positive charge) is attracted more than the rear is repulsed, and the dipole feels a net force toward the line of charge.

He have talked about force and torque, so all that remains from classical mechanics to consider is potential energy. Why should there be potential energy at all? Well, suppose we release the dipole in the diagram above from rest. Clearly it will begin to rotate, which means it will gain kinetic energy. This energy must come from somewhere, and in fact it comes from the work done by the electric field. But the electric field exerts a conservative force, so we can also express it in terms of a potential energy. The change in potential energy due to a conservative force is the negative of the work done by that force. So let's consider the work done by the electric force on the charges of the dipole as the dipole rotates (note, there is no net force on the dipole as a whole, so the movement of its center of mass doesn't change the potential energy).

Figure 1.4.5 – Potential Energy Change for a Rotating Dipole



Only the displacement of each charge along the direction of the force (which is parallel to the electric field) counts toward the work done. The force on each charge has a magnitude of qE, and the force acts in the direction of the displacement for both charges (remember the force on the negative charge is the opposite direction of the field), which makes the work done positive, and the change in potential energy negative. For both charges charge we therefore have:

$$\Delta U_{dipole} = -2W\,(A \rightarrow B) = -2F\Delta l = -2\,(qE)\left(\frac{d}{2}\cos\theta\right) = -\,(qd)\,E\cos\theta \tag{1.4.4}$$

While the torque evoked the idea of a cross product, the potential energy screams out dot product. If we follow a typical convention and define the zero potential energy exist in the configuration when the dipole is perpendicular to the field, then we have:

$$U_{dipole} = -\vec{p}\cdot\vec{E} \tag{1.4.5}$$

Note that the energy is a minimum when the dipole moment aligns with the external electric field.

The Dipole Field

If we are to treat dipoles as "packages," then we have to stop looking at them so closely. Okay, so what do they look like if we look at them from far away? Well, to determine this, we need to look at them as two charges (sigh, again), and look at the field they create at a distance *much greater* than their separation, d. As with the results above, we want our final answer in terms of the dipole moment, not the charge and separation.

We'll start with a simple part of the field – along the axis defined by the dipole moment. Clearly the field points out of the dipole on one end, and into it on the other, following the direction of the dipole moment. For this calculation, we will place both charges on the $x-$axis, equal distances from the origin, with the positive charge on the positive side of the origin. Along the axis a distance r from the origin (on the positive side of the origin), the field of the negative charge is a bit weaker than the field of the positive charge, as it is farther away by a distance d. The exact field at this point is:

$$\overrightarrow{E} = \frac{k\left(-q\right)}{\left(r+\frac{d}{2}\right)^2}\;\hat{i} + \frac{k\left(+q\right)}{\left(r-\frac{d}{2}\right)^2}\;\hat{i} \tag{1.4.6}$$

If we get a common denominator and do a bunch of algebra, we wind up with this:

$$\overrightarrow{E} = \frac{32k\;q\;d\;r}{\left(4r^2-d^2\right)^2}\;\hat{i} \tag{1.4.7}$$

Now let's rearrange things a bit. First, we'll replace the quantity qd with the magnitude of the electric dipole moment, p. Next, we'll divide both the numerator and denominator by r^4:

$$\overrightarrow{E} = \frac{32k\;p\;r^{-3}}{\left(4-\frac{d^2}{r^2}\right)^2}\;\hat{i} \tag{1.4.8}$$

Up to now, we have made no approximations. So finally we invoke what we stated at the outset – let's look at positions that are very far from the dipole compared to the separation of the charges. With this assumption, the ratio of d to r is very small, and the square of that ratio is even smaller, so we treat it as negligible, giving:

$$\overrightarrow{E} = 2k\frac{p}{r^3}\hat{i} \tag{1.4.9}$$

We find the interesting result that while the field of a monopole (point charge) falls-off as $1/r^2$, the field of a dipole (at least along the axis of the dipole moment, and as it turns out, everywhere else – see below) falls off faster – as $1/r^3$. This actually should not surprise us, if we take another look at the dipole field in Figure 1.2.2. We know that the field gets weaker as the field lines diverge from each other. The field lines of a monopole emanate straight out of the charge, and diverge at a constant angle. Along the axis of the dipole, we can see that the field lines *bend away from* the axis, which means they diverge faster than the monopole case, so the field gets weaker faster.

It is a little bit more complicated to work out the field of the dipole off the axis, and we won't go into the details of this derivation, but the final result is:

$$\overrightarrow{E}_{dipole} = \frac{k\left[3\left(\overrightarrow{p}\cdot\hat{r}\right)\hat{r}-\overrightarrow{p}\right]}{r^3} \tag{1.4.10}$$

Example 1.4.2

Show that the general formula for the dipole field yields the proper field along the axis of the dipole.

Solution

Along the axis of the dipole, the position vector (and therefore the position unit vector) for the field is parallel to the dipole vector. If the position of the field is on the positive side of the dipole, then the position unit vector points in the same direction, and:

$$\overrightarrow{p}\cdot\hat{r}=\left|\overrightarrow{p}\right|\;\left|\hat{r}\right|\;\cos 0^{o}=p\quad\Rightarrow\quad 3\left(\overrightarrow{p}\cdot\hat{r}\right)\hat{r}=3p\hat{r}=3\overrightarrow{p}$$

Plugging this back into the general formula gives the result on the axis.

1.5: Conductors

A New Constant

The constant k that we introduced in Coulomb's law, and is present in all of the solutions above, was convenient in the context for which it was introduced, but we will introduce a new constant here that we will use hereafter. There is no fundamental difference in using either of these constants – mostly the change is for cosmetic reasons. Without further ado, presenting the *permittivity of free space*:

$$\epsilon_o \equiv \frac{1}{4\pi k} = 8.85 \times 10^{-12} \frac{C^2}{Nm^2}$$

Rewriting the magnitudes of the coulomb field and the fields of the infinite distributions in Section 1.3 in terms of this constant, we have:

$$E_{point}(r) = \frac{q}{4\pi\epsilon_o r^2} \tag{1.5.1}$$

$$E_{line}(r) = \frac{\lambda}{2\pi\epsilon_o r} \tag{1.5.2}$$

$$E_{plane} = \frac{\sigma}{2\epsilon_o} \tag{1.5.3}$$

We will make extensive use of these formulas from this point on. We will also discard the use of k in favor of ϵ_o in all forthcoming applications.

Electrostatics

Before we discuss the effects of conductors on electric fields, it is essential that we make clear that we are proceeding under the following restriction: We are talking about *electrostatics*. This means that the charges present are in a state of static equilibrium – they are not moving, nor are they accelerating. We have already said that conductors allow for charges to flow freely, so how can these two things be reconciled?

If a conductor suddenly finds itself in the presence of an electric field, then the charges on that conductor will start to move as a result of the new force. As the charges move, the fields (which are affected by the placement of these charges) also change. The charges continue accelerating until the field contributions of the displaced charges cancel the external field. With zero net field, the charges no longer accelerate, and we will assume that their kinetic energy dissipates such that they also come to rest.

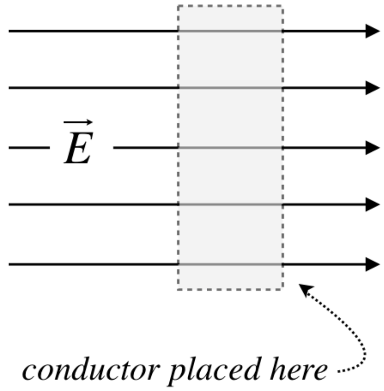
The conclusions we will draw here do not apply to the period of time when the charges are still moving around – we will only be considering the case when the charges finally reach an equilibrium electrostatic state.

Inside Conductors

We characterized electric fields as "signals" sent out by electric charges. This is of course a model (as is all of physics), and this model does not make exceptions that allow for matter to "block" this signal. The only way to affect the field at a point in space caused by one charge is to introduce a field from another charge, such that the two fields superpose.

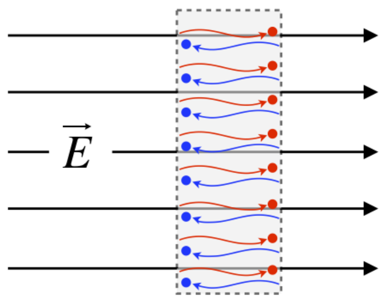
Let's now consider what happens when we suddenly introduce a uniform external electric field to a rectangular conducting slab. We represent uniform electric fields with parallel, equally-spaced field lines, and as we just said above, these field lines are not interrupted by matter, so the situation looks something like this:

Figure 1.5.1 – Conductor Introduced to Field



While the conductor is neutrally-charged, it is not without electrical charge – it is made of atoms, which are comprised of protons and electrons. This charge is (by definition of the conductor) free to move. This means that the positive charges in the conductor in the diagram (depicted as red) are pushed to the right by the electric field, while the negative charges (depicted as blue) are pulled to the left.

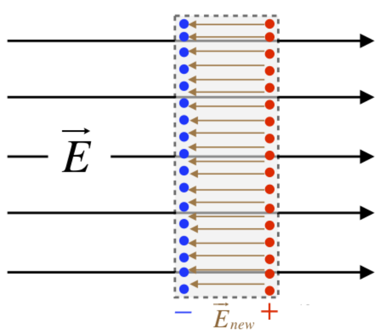
Figure 1.5.2 – Charges Migrate Because of Applied Field



[*Okay, so technically only the electrons – the negatively-charged particles – are free to move, or this conductor would not be a solid. This is a distinction that will not be at all important to us, because when negatively-charged electrons vacate a region, they leave an excess of positively-charged protons behind, which is completely equivalent to the positive charge moving into that electron-vacated region.*]

The effect of this migration of charge is the creation of two planes of charge, one on each side of the slab (we will always assume that the field is not strong enough to pull the electrons off the surface of the metal). But this separation of charge itself has consequences with regard to the field. In particular, *within the metal* a new uniform field starts to develop. This field points away from the positive charges toward the negative charges, which opposes the direction of the external field.

Figure 1.5.3 – Internal Field Induced by Displaced Charges



When the field induced within the conductor is superposed with the applied field, the result is a weaker field within the conductor. The question is, how much weaker does it become? Suppose it only becomes a little weaker. That would mean that there is still some net field pointing left-to-right. But if this is the case, then more charge will migrate, making the induced field that points left even stronger, making the superposed field even weaker. In other words, we don't reach electrostatics until enough charge has moved to make the superposed internal field vanish. We therefore get the following remarkably general conclusion:

Electric fields vanish within conductors when the charges are static.

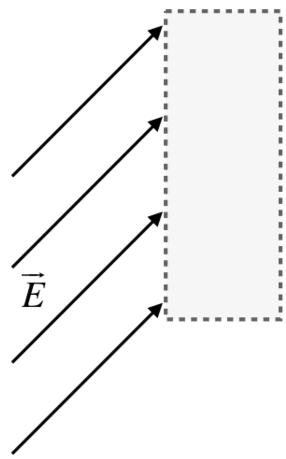
Alert

*It's important to remember that this rule applies **within the metal itself**. If a conductor is hollow, the interior space does not qualify as being "within" the conductor, and the field is not necessarily identically zero.*

At the Surface of Conductors

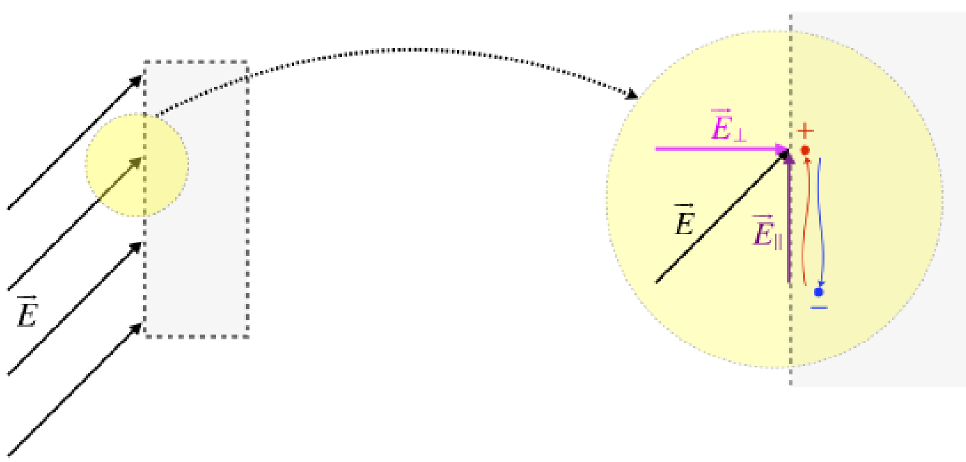
The example above assumed that the uniform field made a right angle with the surface of the conductor, so let's generalize this a bit by allowing the field to make a different angle.

Figure 1.5.4 – Conductor Introduced to Field at an Angle



We handle this case as we would with any vectors – by breaking them into components. The component perpendicular to surface will affect the charges in the conductor exactly as described above – they will migrate to the two surfaces until the horizontal part of the field within the conductor vanishes. But what about the component parallel to the surface? This will only affect the charges at the surface, but again it will cause a migration. Here is a blown-up depiction of what is going on:

Figure 1.5.5 – Surface Migration of Charges

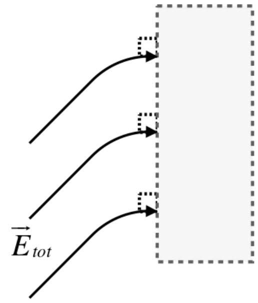


As before, these displaced charges result in an induced field, which points from the positive charges toward the negative charges, opposing the parallel component of the external field. The charges keep migrating until the superposition of the induced field and the parallel component of the external field cancel each other. Unlike the case of inside the conductor, the whole field does not vanish in this case – just the parallel component does. The result is another remarkably general conclusion:

Electric field lines strike conducting surfaces at right angles.

The migrated charges create fields that extend out into the nearby space, and this causes the total electric field to adjust "smoothly" into the perpendicular surface strike:

Figure 1.5.6 – Field Lines Land Orthogonally



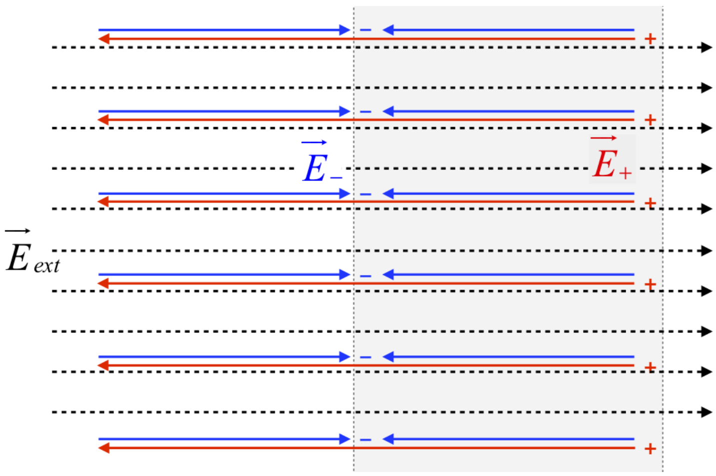
Alert

It might appear from our specific example that this result only applies to flat conducting surfaces, but it is more general than that. The charges cannot remain static at any point on the surface of any conductor if the electric field at that point has a parallel component. So charges on a conductor rearrange themselves such that the total electric field lands at right angles to the conducting surface regardless of its shape.

Field Strength at a Conducting Surface

Now that we know about the direction of an electric field at the surface of a conductor, let's use a simple model to have a look at the magnitude. We return to our example of a conducting slab with a perpendicular field (Figure 1.5.1), but here we will assume that the slab is infinite in extent (up/down and into/out of the page), with a finite thickness. When the charges migrate for this slab, they form two infinite planes of charge, for which we have computed the electric field strength. We found that these fields are uniform and are proportional to the charge density (Equation 1.3.22). Note that for every electron that migrates, it leaves behind a proton, which means that the charge density of both planes of charge are equal in magnitude (but have opposite signs). So our physical model has three separate electric fields that superpose: The applied electric field, the induced electric field of the negative plane of charge, and the induced electric field of the positive plane of charge. Putting these all together looks like this:

Figure 1.5.7 – Three Source Fields Near a Conducting Slab



The shaded area is the slab, and we are focused on the field to each side of the left surface. The external field has drawn the negative charges to this surface, while pushing positive charges to the other surface (or rather, leaving them behind). The magnitudes of the fields $\overrightarrow{E}_-$ and $\overrightarrow{E}_+$ are equal, and are, in terms of the charge density at each surface:

$$|E_-| = |E_+| = \frac{|\sigma|}{2\epsilon_o} \tag{1.5.4}$$

The fact that the field $\overrightarrow{E}_-$ points toward the negative charges and $\overrightarrow{E}_+$ points away from the positive charges means that the two fields point in the same direction inside the conductor, resulting in a total induced field inside the conductor with a magnitude of $\frac{\sigma}{\epsilon_o}$. But as we already know, the field inside the conductor vanishes, so we must have that enough charge must migrate such that:

$$|E_{ext}| = |E_-| + |E_+| = \frac{\sigma}{\epsilon_o} \tag{1.5.5}$$

Now let's look outside the left surface of the slab. In that region, the two induced fields are pointing in *opposite* directions. They are equal in magnitude, so they simply cancel each other, giving a total field strength equal to just the external field. Therefore we have for the total electric field outside the surface of the conductor:

$$|E_{tot}| = \frac{\sigma}{\epsilon_o} \tag{1.5.6}$$

While this appears to be a very specific result, we will shortly show that it applies to every conducting surface (not just an infinite slab). In such cases, the charge density can be different at different points on the surface of the conductor, but whatever the density happens to be *at a specific point* gives us the electric field *at that point* according to this equation. Given that we also know that the electric field is perpendicular to the surface at this point, we know everything there is to know about the field there. This is a strikingly simple and powerful result that we will use over and over.

1.6: Gauss's Law

The Density of Field Lines

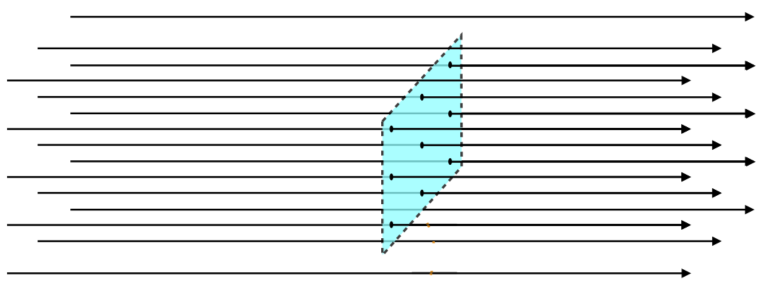
We have said that the strength of an electric field in a given region is represented by the density of the electric field lines in that region. First, it is important to note that technically there is an *infinite* number of field lines. When we draw a diagram with a finite number of field lines, it is easy to tell where they are closer together and farther apart. But even when we extend to an infinite number the idea of line *density* holds. We now seek to quantify this measure somewhat.

Whenever we talk about the density of some quantity, it always comes as some kind of ratio:

$$\text{density} \equiv \frac{\text{amount of something}}{\text{volume or area or length that it occupies}} \tag{1.6.1}$$

If we want to measure a density of infinitely-long field lines, we cannot do it using a ratio with a volume (we cannot contain them in a volume). Instead, we do it with an area, by counting the number of field lines that "pierce" a surface, and divide it by the area of that surface. So for example:

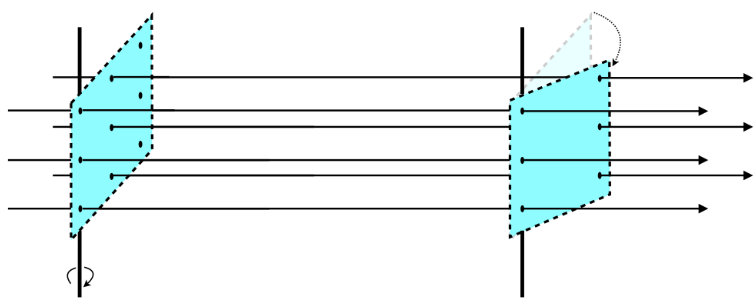
Figure 1.6.1 – Counting Surface Piercings By Field Lines



So in the case above, we would divide 9 piercings by the area of the rectangular surface to get the field line density. If we take the same rectangular surface elsewhere, and the field pierces it 18 times, we would conclude that the field is twice as strong there.

There is one problem with this scheme. We haven't specified the *orientation* of the surface relative to the field lines. For example, in the case above, if we rotate the rectangle slightly, we get a different number of piercings:

Figure 1.6.2 – Dependence of Surface Piercings on Surface Orientation



There is one unique way to define the density, and it results in the *maximum* number of piercings – define the density only in terms of a surface through which the piercings are *perpendicular*. So we define the field line density (which we equate to field strength) as follows:

$$\text{field line density} \equiv \frac{\text{number of field lines piercing a surface perpendicularly}}{\text{area of that surface}}$$

Notice that since there are infinitely-many field lines, we can define this surface to be as small as we like. We can then use an infinitesimally-small surface to probe various positions in space to determine the field strength at every point, provided we rotate it to be perpendicular to the field.

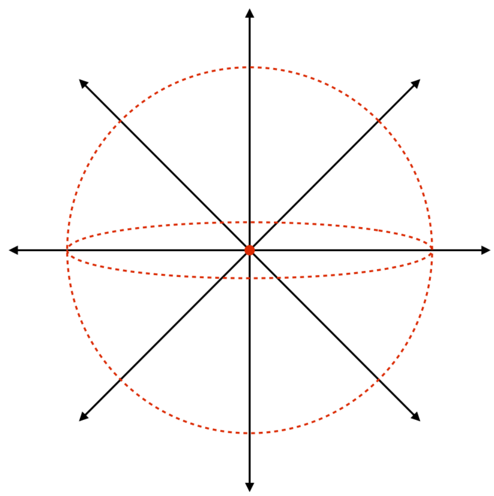
The Role of Charge

From the equation of the coulomb field, it's clear that the strength of the field is directly proportional to the amount of charge. That is, if we don't change the *distribution* of the charge, but just double the total amount everywhere (i.e. double the charge density everywhere it exists), then the field will look exactly like before, but it will be twice as strong. But we also measure field strength as density of field lines. If we use the same surface twice at the same point in space, once with the original charge, and once with double the charge, the density of field lines must double. Given that the field lines should not change direction, the only way that the field line density can double in this case is if the total number of field lines doubles. We therefore conclude that the *number of field lines due to a collection of charge is proportional to the total amount of charge*. For now we will write it this way:

$$Q = charge = (constant) \cdot (number\ of\ field\ lines\ coming\ out\ of\ it) = \mathcal{C} \cdot \mathcal{N} \tag{1.6.2}$$

There is a clever way that we can determine what this constant of proportionality is. It involves choosing a single positive point charge, and enclosing it at the center of of a spherical surface.

Figure 1.6.3 – Point Charge Enclosed by Two Spherical Surfaces



These field lines pierce the spherical surface *perpendicularly*, and the density of field lines is the same regardless of where you look on the surface of the sphere. Therefore the field line density is the number of field lines piercing the surface divided by the total surface area of the sphere ($4\pi r^2$). And since this surface completely encloses the charge, the number of field lines piercing the surface is the total number of field lines produced by that charge ($\mathcal{N}$). The field line density at this spherical surface is therfore:

$$density\ of\ field\ lines = \frac{\mathcal{N}}{4\pi r^2} \tag{1.6.3}$$

But the density of field lines is the electric field strength, which we happen to know at the surface of this sphere (a distance r from the point charge), from coulomb:

$$E(r) = \frac{Q}{4\pi\epsilon_o r^2} \tag{1.6.4}$$

Putting these last two equations tells us the number of field lines coming out of a charge Q:

$$\mathcal{N} = \frac{Q}{\epsilon_o} \tag{1.6.5}$$

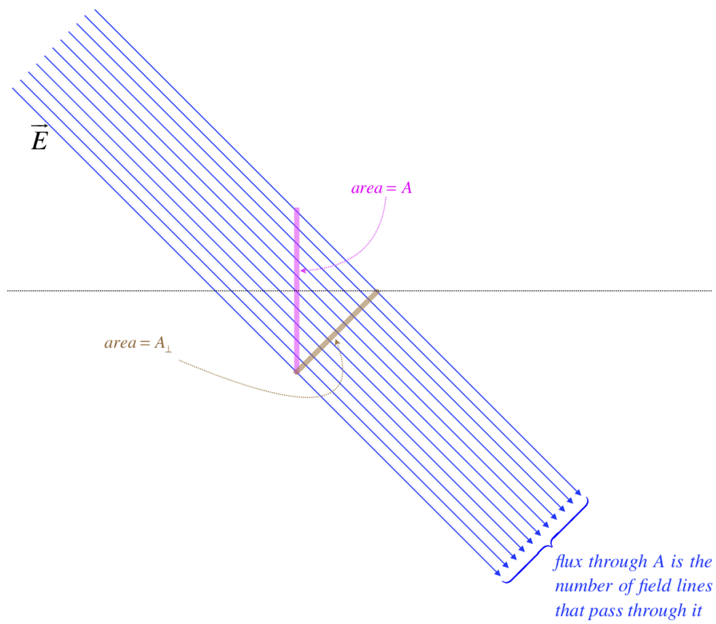
So we see that the constant of proportionality $\mathcal{C}$ from Equation 1.6.2 is simply our old friend ϵ_o. Notice that it didn't matter what the radius of the sphere was. The number of field lines piercing the spherical surface is just however many are coming out of the charge Q, and to get that number we just divide Q by ϵ_o. In fact, it didn't even depend upon the fact that we used a sphere at all. So long as we choose a *closed surface* that encloses the charge, the number of field line piercings will be the same – it depends only upon the amount of charge. With another surface, the field line density will differ from one place to the other on the surface, but the number of piercings won't change.

Electric Field Flux

We have been dancing around the issue of there being an infinite number of field lines. That is, the number of field lines coming from a charge Q is not really *equal* to $\frac{Q}{\epsilon_o}$, any more than the electric field strength is really *equal* to the field line density. This is evident (if not conceptually) from the fact that the units don't match. So we need to drag our understandable-but-imprecise model into the world of more rigorous mathematics.

The link between these two worlds is the concept of *electric field flux* through a surface. In our description above, this is simply the number of field line piercings. But as we said, this is actually infinite, so we need the more mathematically rigorous definition of this term. We can get to this by relating the total number of piercings to the field line density.

Figure 1.6.4 – Definition of Electric Field Flux



The number of field lines that pass through the pink surface with area A is equal to the number of field lines that pass through the brown surface with area $A_\perp$. The number of field lines divided by $A_\perp$ is the electric field strength, so if we call the electric field flux (i.e. the number of penetrating field lines) "Φ_E," then we have:

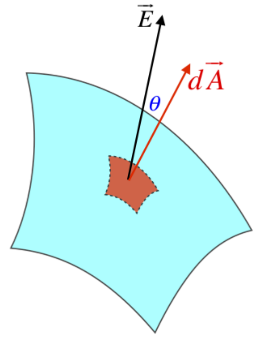
$$E = \frac{\Phi_E}{A_\perp} \tag{1.6.6}$$

We can write this in terms of the original (pink) surface if we know the angle that the field makes with it. It is standard to measure this angle with a line perpendicular to that area (in the diagram above, that would be with the horizontal line shown). Doing the geometry reveals that the areas are related through the cosine of the angle, giving:

$$A_\perp = A\cos\theta \quad \Rightarrow \quad \Phi_E = E\;A\cos\theta \tag{1.6.7}$$

Now we are not always fortunate enough to have situations where we have a nice, flat surface and a uniform electric field with which to calculate the flux. When that is the case, we can only do small bits of flux at a time, and then add the small bits together to get the total flux.

Figure 1.6.5 – Differential Flux



The differential area vector is defined with a magnitude equal to its area (which is infinitesimally small) and a direction that is perpendicular to the surface at that point on the surface. The electric field is also a vector, so when we compute the (tiny) flux through this area and we have to include the cosine of the angle, it's not surprising that we describe it with a dot product:

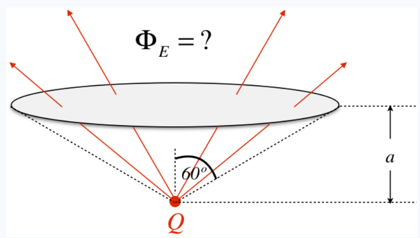
$$d\Phi_E = E\,dA\cos\theta = \overrightarrow{E}\cdot d\overrightarrow{A} \tag{1.6.8}$$

If we want to compute the flux over a finite surface, we have to add all of these infinitesimal contributions:

$$\Phi_E\left(surface\right) = \int\limits_{surface} \overrightarrow{E}\cdot d\overrightarrow{A} \tag{1.6.9}$$

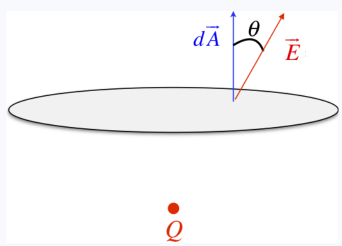
Example 1.6.1

Find the flux of a point charge Q lying on the axis of a flat circular surface a distance a from the charge. The radius of the circular surface is such that a straight line joining the point charge and the edge of the surface makes a 60^o angle with the axis (see the diagram below). Compute the electric flux through the surface.

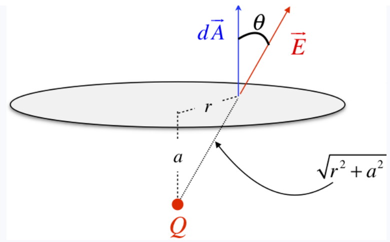


Solution

We begin by computing the radius of the circular surface. The 30/60/90 triangle with the side opposite the 30^o angle equal to a must have a hypotenuse equal to $2a$, which gives it a longer side (the radius of the circular surface) equal to $\sqrt{3}a$. We next need to come up with the flux through an arbitrary infinitesimal piece of the circle. Note that the electric field is not perpendicular to this surface everywhere, since it comes radially out of the point charge. At an arbitrary point on the circle, it looks like this:



The simplest coordinate system to use here is cylindrical, with the z-axis being the line passing through the center of the circle and the charge. The distance out to the arbitrary position from the center is the cylindrical coordinate r.



Trigonometry gives us the angle θ in terms of a and r:

$$\cos\theta = \frac{a}{\sqrt{r^2+a^2}}$$

The magnitude of the electric field comes from coulomb:

$$E = \frac{Q}{4\pi\epsilon_o\left(r^2+a^2\right)}$$

A tiny section of area at the point in question in cylindrical coordinates is the usual (arclength multiplied by radial length):

$$dA = r\ dr\ d\phi$$

Now put all of this together to compute the flux integral:

$$\Phi_E = \int \vec{E}\cdot d\vec{A} = \int E\ dA\cos\theta = \int \frac{Q}{4\pi\epsilon_o\left(r^2+a^2\right)}\left(r\ dr\ d\phi\right)\frac{a}{\sqrt{r^2+a^2}} = \frac{Qa}{4\pi\epsilon_o}\int_0^{2\pi} d\phi \int_0^{\sqrt{3}a} \frac{r}{\left(r^2+a^2\right)^{\frac{3}{2}}}dr$$

Integrating gives:

$$\Phi_E = \frac{Qa}{4\pi\epsilon_o}(2\pi)\left[\frac{-1}{\sqrt{r^2+a^2}}\right]_0^{\sqrt{3}a} = \boxed{\frac{Q}{4\epsilon_o}}$$

Gauss's Law

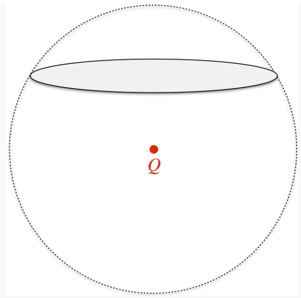
The culmination of everything we have done in this section comes rather quickly by combining our revelations above with the mathematical definition of flux. We found that when we enclose a charge inside a closed surface, the number of piercings of that surface equals the total number of field lines produced by that charge, which we said was $\frac{Q}{\epsilon_o}$. But the total number of piercings is also the flux through that surface. We therefore conclude that the total flux through a closed surface is given by:

$$\oint \vec{E}\cdot d\vec{A} = \frac{Q_{encl}}{\epsilon_o} \tag{1.6.10}$$

The integral with a circle in it is shorthand for "over a closed surface," and the subscript on Q indicates that the charge is enclosed in this surface. The direction of the vector $d\vec{A}$ is *out of the closed volume*. This assures that the integral comes out positive for positive charges, and negative for negative charges. This is known as the integral form of Gauss's law. In the next section we'll look at the applications of this law.

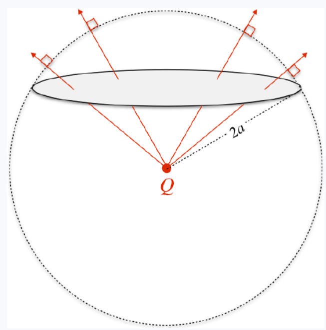
Example 1.6.2

Repeat the flux calculation of the previous example, this time using Gauss's law with a closed sphere centered at the point charge and circumscribed around the surface, as shown in the diagram below.



Solution

We want to know the flux through the circular surface. The top section of the sphere (separated from the rest of the sphere by the circle) has no enclosed charge, so according to Gauss's law, the total flux out of its closed surface is zero. This means that the inward flux (coming through the circle from the charge) equals the outward flux (exiting the sphere in the top section). We can therefore determine the flux through the circle if we can compute the flux through the top section of the spherical surface.



The flux through the top section is easier to compute because the field lines are perpendicular to this surface and has the same magnitude everywhere. The infinitesimal area element in spherical coordinates is $r^2 \sin\theta \, d\theta \, d\phi$*, where in this case* $r = 2a$*, so the flux integral is:*

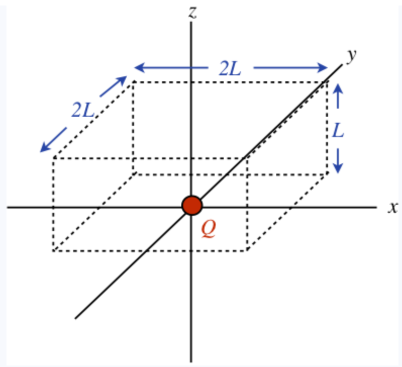
$$\Phi_{circular\ surface} = \Phi_{top\ of\ sphere} = \int \overrightarrow{E} \cdot d\overrightarrow{A} = \int E \, dA \cos 0^o = \int\int \frac{Q}{4\pi\epsilon_o (2a)^2} (2a)^2 \sin\theta \, d\theta \, d\phi$$

The limits of integration are pretty obvious: The azimuthal angle ϕ *goes all the way around (* $0 \to 2\pi$ *), while the polar angle* θ *goes from vertical to the edge of the circle (* $0 \to 60^o$ *). The angular integrals are straightforward, and putting it all together gives the same answer as obtained previously.*

If one happens to know that one quarter of the surface area of a sphere is subtended by a 60^o *solid angle (okay, so maybe not too many people have this information readily available), then the solution comes even faster. According to Gauss's law, the total flux out of the sphere is* $\frac{Q}{\epsilon_o}$*, and since the charge is at the center of the sphere, the flux is uniform over the whole surface, which means that one fourth of the total flux must be going out of the top section, giving the answer immediately.*

Example 1.6.3

A point charge Q lies in the x*-*y *plane. Consider an imaginary block (all vertices are right angles) with dimensions* $2L \times 2L \times L$*, positioned with one of its square faces flat on the* x*-*y *plane, centered at the point charge (see diagram). Through a Herculean display of mathematical prowess, one can show via direct integration that the electric field flux through the top* $2L \times 2L$ *surface comes out to be* $\frac{Q}{6\epsilon_o}$*. Use Gauss's law to achieve the same result in a manner requiring far less effort.*



Solution

if we construct an identical imaginary block below the x-y plane, then the two blocks combined form a single cubical gaussian surface that encloses the charge. The charge is at the exact center of this cube, so there is no reason that any more field lines would pierce one surface of the cube than any other. There are 6 faces to the cube, so one-sixth of the total flux out of the cube passes through top $2L \times 2L$ surface. According to Gauss's law, the total flux escaping the cube is $\frac{Q}{\epsilon_o}$, so we get the same answer as above.

Local (Differential) Form of Gauss's Law

Gauss's law can be cast into another form that can be very useful. There is a theorem from vector calculus that states that the flux integral over a closed surface like we see in Gauss's law can be rewritten as a volume integral over the volume enclosed by that closed surface. It is called the *divergence theorem*:

$$\oint \overrightarrow{E} \cdot d\overrightarrow{A} = \int\limits_{\substack{volume \\ enclosed}} \left(\overrightarrow{\nabla} \cdot \overrightarrow{E}\right) dV \tag{1.6.11}$$

Alert

It should be noted that this theorem is sometimes expressed a bit differently, particularly in mathematics texts. Namely, the differential area vector is broken into two parts – the magnitude dA and the direction $\widehat{n}$ (a unit vector pointing perpendicularly out of the enclosed surface at the position where the area element is located), giving:

$$\oint \overrightarrow{E} \cdot \widehat{n}\, dA = \int\limits_{\substack{volume \\ enclosed}} \left(\overrightarrow{\nabla} \cdot \overrightarrow{E}\right) dV$$

If we now apply this to Gauss's law, we get:

$$\frac{Q_{encl}}{\epsilon_o} = \int \left(\overrightarrow{\nabla} \cdot \overrightarrow{E}\right) dV \tag{1.6.12}$$

But we can write the charge enclosed in a volume in terms of the charge density ρ within that volume:

$$Q_{encl} = \int \rho\, dV \tag{1.6.13}$$

Comparing these last two equations suggests the following result:

$$\overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho}{\epsilon_o} \tag{1.6.14}$$

This is simply a restatement of Gauss's law, and is known as the *local (or differential) form* of that law.

This version can be used to solve the same kinds of problems as the integral form (though with a slightly different method), and is especially useful when the field's positional dependence is known (divergences are easier to compute than integrals).

Divergence Formulas

Now that it is clear that the divergence of the electric field is an important thing to be able to calculate, it is useful to point out some useful formulas for the divergence operation. The first of these is usable under all conditions:

Cartesian Coordinates

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}\left(x,\ y,\ z\right)=\frac{\partial}{\partial x}E_x+\frac{\partial}{\partial y}E_y+\frac{\partial}{\partial z}E_z \tag{1.6.15}$$

The other two coordinate systems we will encounter frequently are cylindrical and spherical coordinates. In terms of these variables, the divergence operation is significantly more complicated, *unless* there is a radial symmetry. That is, if the vector field points depends only upon the distance from a fixed axis (in the case of cylindrical coordinates), or upon the distance from a fixed point (in the case of spherical coordinates), the divergence operation is fairly straightforward:

Cylindrical Coordinates

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}\left(r,\ \not\phi,\ \not z\right)=\frac{1}{r}\frac{\partial}{\partial r}\left(rE_r\right) \tag{1.6.16}$$

Spherical Coordinates

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}\left(r,\ \not\theta,\ \not\phi\right)=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2E_r\right) \tag{1.6.17}$$

Example 1.6.4

An electric field in a region of space near the origin can be expressed as:

$$E\left(x,y,z\right)=\alpha\ x\ \hat{i}+\beta\left(y+y_o\right)^2\ \hat{j}$$

Show that if there is no charge at the origin, then the electric field there is equal to $\dfrac{\alpha^2}{4\beta}\ \hat{j}$.

Solution

By Gauss's law, we have:

$$\frac{\rho}{\epsilon_o}=\overrightarrow{\nabla}\cdot\overrightarrow{E}=\overrightarrow{\nabla}\cdot\left[\alpha x\ \hat{i}+\beta\left(y+y_o\right)^2\ \hat{j}\right]=\frac{\partial}{\partial x}\left[\alpha x\right]+\frac{\partial}{\partial y}\left[\beta\left(y+y_o\right)^2\right]=\alpha+2\beta\left(y+y_o\right)$$

Zero charge at the origin means that the charge density vanishes there, so setting the density equal to zero at ($x=y=z=0$ *) gives:*

$$0=\alpha+2\beta y_o\quad\Rightarrow\quad y_o=-\frac{\alpha}{2\beta}\quad\Rightarrow\quad\overrightarrow{E}\left(x,y,z\right)=\alpha\ x\ \hat{i}+\beta\left(y-\frac{\alpha}{2\beta}\right)^2\ \hat{j}$$

Plugging in ($x=y=z=0$ *) gives the desired answer.*

1.7: Using Gauss's Law

Symmetry Avoids Integrals

The great irony of Gauss's law is that the surface integral looks incredibly daunting, but this law is only really useful because no integration actually needs to be performed. As we will see, we will be able to use this law to compute electric fields of distributions of charge in cases where some degree of symmetry is present. The basic approach is this: Construct an imaginary closed surface (called a *gaussian surface*) around some collection of charge, then apply Gauss's law for that surface to determine the electric field at that surface. This is a rather vague description, and glosses over a lot of important details, which we will learn through several examples.

There are two ingredients to the symmetry that need to be present to make using Gauss's law so powerful:

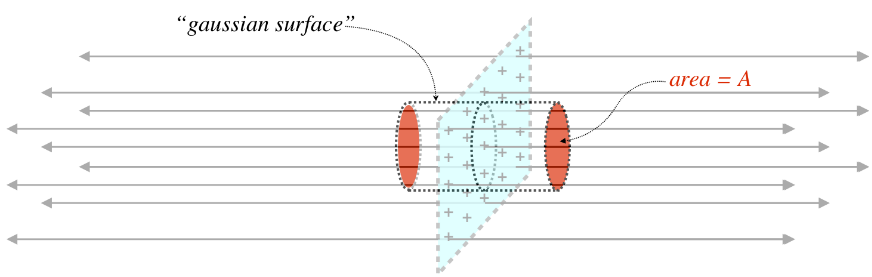
1. A gaussian surface must exist where the electric field is either parallel or perpendicular to the surface vector. This makes the cosines in all the dot products equal to simply zero or one.
2. The electric field that passes through the parts of the gaussian surface where the flux is non-zero has a constant magnitude.

These two conditions allow us to avoid an integral entirely, because the $cos\theta$ in the integral goes away, and the electric field magnitude can be taken out of the integral, leaving only an integral of dA, which is just the area of the surface. Then applying Gauss's law is simple.

Field of an Infinite Plane of Charge

This is a problem we have already solved (Equation 1.3.22). We did it by computing the field of a disk of charge on the axis, then taking the limit as the radius of the disk goes to infinity. That was a lot of math! Let's see how we can do it with Gauss's law. It's clear that an infinite plane of positive charge must create a field that points away from, and perpendicular to, the plane in both directions. Let's choose as our gaussian surface a cylinder whose axis is perpendicular to the plane of charge, with a cross-sectional area A.

Figure 1.7.1 – Gaussian Surface for a Plane of Charge



Alert

One of the hardest ideas to grasp for students learning for the first time about how to use Gauss's law seems to be the idea that the gaussian surface is something we construct ourselves as a problem-solving tool. There is no actual surface present, nor is there a specific unique surface that must be used. To make the solution as simple as possible, the surface should have the two properties given above, and the trick to these problems is conceiving of a surface that does this.

We note that the electric field only passes through the ends of the cylinder, which means that there is no flux through the sides. Also, the field that passes through the ends is parallel to the area vector, so $\cos\theta=1$ everywhere on that surface. The electric field strength is the same value everywhere on the surface, so it can be pulled out of the integral, which then gives simply the area of the end of the cylinder. The flux is out (positive) at both ends and equal, so they provide equal contributions to the total flux. The total flux out of the cylinder then is simply:

$$\Phi_E = \cancelto{0}{\Phi_E\left(sides\right)} + \Phi_E\left(left\ end\right) + \Phi_E\left(right\ end\right) \quad\Rightarrow\quad \Phi_E = 2EA \tag{1.7.1}$$

Now we apply Gauss's law. The amount of charge enclosed in this cylinder is the surface density of the charge multiplied by the area cut out of the plane by the cylinder (like a cookie-cutter), which is clearly equal to A, the area of the ends of the cylinder. Applying Gauss's law gives:

$$\Phi_E = \frac{Q_{encl}}{\epsilon_o} \quad \Rightarrow \quad 2EA = \frac{\sigma A}{\epsilon_o} \quad \Rightarrow \quad E = \frac{\sigma}{2\epsilon_o} \tag{1.7.2}$$

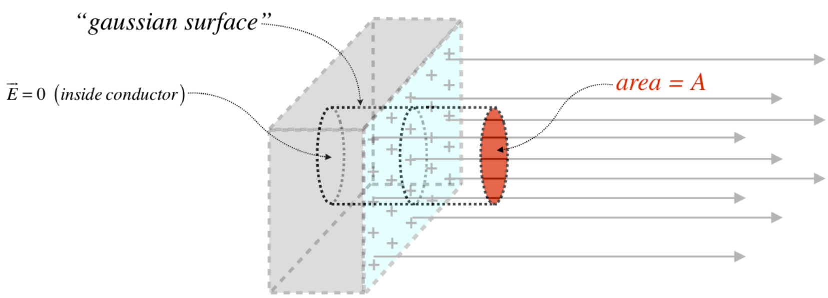
This is exactly the answer we got before! Notice that the final answer comes out to be independent of the length of the cylinder, which means that the field is uniform, and it comes out to be independent of the area of the cylinder as well.

One might well ask, "What if the cylinder didn't have straight sides? That is, what if it bulged in the middle, causing it to enclose more charge? Won't this give a different answer?" The answer is that if the sides of the cylinder aren't straight, then the electric field will pierce the gaussian surface through the sides as well as the ends. The flux through the ends would be the same as before, and the additional flux through the sides would account for the additional enclosed charge.

Field Outside an Infinite Charged Conducting Plane

We have already solved this problem as well (Equation 1.5.6). Solving it with Gauss's law is almost identical to the case above, with one exception: We don't know what the field looks like on both sides of the conductor – we only know that one side of it is charged. But that's fine, because this time we choose our gaussian cylinder so that one end surface is outside the conductor, and the other is *inside the metal*.

Figure 1.7.2 – Gaussian Surface for a Conducting Plane of Charge



The computation follows exactly as before, with one exception: We know that conductors have zero electric field inside the metal (we are assuming electrostatics here), so there is no electric field flux through that end of the cylinder. The enclosed charge is the same as before, so we get:

$$\Phi_E = \frac{Q_{encl}}{\epsilon_o} \quad \Rightarrow \quad EA = \frac{\sigma A}{\epsilon_o} \quad \Rightarrow \quad E = \frac{\sigma}{\epsilon_o} \tag{1.7.3}$$

Once again, the same answer that we got previously. But there is additional value in this solution that we didn't have before. In our previous approach to this, we made some specific assumptions about the shape of the conducting slab. With Gauss's law, we can even work with a *curved* surface, for the following reason: When a surface is curved, that curvature is only noticeable when a sufficient amount of that surface is taken into account (e.g. the Earth's surface appears to be flat until you get far enough away from it). In this gauss's law approach, we can make the cross-sectional area of the cylinder as arbitrarily small as we like, and the answer doesn't change. As soon as we make the cross-sectional area "small enough" that the curved conducting surface is effectively flat (i.e. the electric field is constant over the entire end surface of the cylinder), then the answer obtained applies. This means that this answer applies at *every conducting surface*, if the density is evaluated at a specific position on the surface. In other words, if the charge density on the surface of a conductor at position x is $\sigma\left(x\right)$, then the electric field magnitude at that same position in space is:

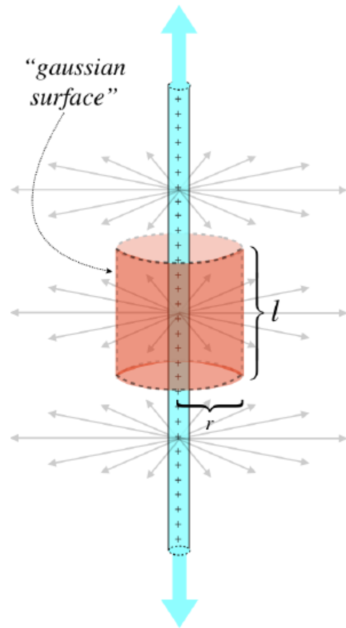
$$E\left(x\right) = \frac{\sigma\left(x\right)}{\epsilon_o} \tag{1.7.4}$$

An as we already found, the field is perpendicular to the conducting surface at that point).

Field of an Infinite Line of Charge

Yet another problem we have already solved! As we will see, this one is different from the previous two, in that the field will end up depending upon the dimensions of our gaussian surface. This gives us a field that is not uniform (which it isn't!). Once again, the trick is to define a gaussian surface where the field lines pass through parts of it at right angles, and other parts not at all. The obvious choice is therefore a cylinder.

Figure 1.7.3 – Gaussian Surface for an Infinite Line of Charge



We know from symmetry arguments we have already made in the past that the field points radially outward from the line, which means that the field lines don't pass through the ends of the cylinder, contributing nothing to the total flux. Though the curved surface of the cylinder, the electric field is perpendicular everywhere, and since the cylinder is centered at the line of charge, the field strength is the same everywhere. The total flux is therefore the electric field strength at the cylinder wall multiplied by its area:

$$\Phi_E = \cancelto{0}{\Phi_E\left(top\right)} + \cancelto{0}{\Phi_E\left(bottom\right)} + \Phi_E\left(sides\right) \quad \Rightarrow \quad \Phi_E = EA = 2\pi r l E \tag{1.7.5}$$

The enclosed charge is the charge contained between the two ends of the cylinder, which is the linear charge density multiplied by the length of the segment, which is the length of the cylinder. Applying Gauss's law therefore gives:

$$\Phi_E = \frac{Q_{encl}}{\epsilon_o} \quad \Rightarrow \quad 2\pi r l E = \frac{\lambda\, l}{\epsilon_o} \quad \Rightarrow \quad E = \frac{\lambda}{2\pi\epsilon_o\, r} \tag{1.7.6}$$

Again this is in agreement with the answer previously obtained (Equation 1.3.21).

Fields Within Charge Distributions

The reader should not get the impression that electric fields only exist *outside* of charge distributions, though so far every example has been of this variety. Indeed Gauss's law is very useful for finding fields *within* charge distributions, and the process is really no different from what is outlined above.

Consider the case of a sphere of charge with a uniform density ρ and a radius R. We can use Gauss's law to compute the electric field at points within the region of the charge distribution ($r < R$), as well as outside the sphere ($r > R$). The latter calculation is as simple as those above – the field has spherical symmetry (is radially outward), so we choose a spherical gaussian surface (through which the field will pass orthogonally, and on which the field strength is constant), giving:

$$EA_{sphere} = \frac{Q_{encl}}{\epsilon_o} \quad \Rightarrow \quad E\left(r\right) = \frac{Q_{encl}}{4\pi\epsilon_o r^2} \tag{1.7.7}$$

Yes, the field looks exactly like that of a point charge! This will be true for the empty space outside of *all* spherically symmetric charge distributions, even if the charge density varies with respect to r. As we are not given the value of Q, we are not yet finished with this problem. The density is constant, so the total charge is just the density multiplied by the volume of the charge. *Note that this is not the volume of our gaussian surface*, which resides outside the sphere, so:

$$E(r)=\frac{\rho V}{4\pi\epsilon_o r^2}=\frac{\rho\frac{4}{3}\pi R^3}{4\pi\epsilon_o r^2}=\frac{\rho R^3}{3\epsilon_o r^2} \tag{1.7.8}$$

Okay, so what about *within* the charge distribution? The solution is performed in precisely the same way, except that now the spherical gaussian surface has a radius r that is less than R. So how does this change the answer? Well, there is *less charge enclosed* than in the previous case. Specifically, this time the entire gaussian surface is filled with charge. Plugging in this new, smaller volume gives:

$$E(r)=\frac{\rho V}{4\pi\epsilon_o r^2}=\frac{\rho\frac{4}{3}\pi r^3}{4\pi\epsilon_o r^2}=\frac{\rho}{3\epsilon_o}r \tag{1.7.9}$$

Rather than getting weaker with an inverse-square dependence as it gets farther from the center, this field actually gets *stronger* linearly. This happens until r reaches the outer surface of the sphere of charge, then after that it follows the point-charge-like inverse-square weakening behavior.

Alert

Whenever one solves a problem that includes multiple regions like this one (one region being inside the charge, and the other outside the charge), it is a good idea to check to make sure that the field is continuous at the boundary. Indeed, in this case, if we plug $r=R$ into both the interior and exterior solutions, we get the same result. We will see that this is also sometimes used as a condition that we impose to help us solve the problem.

Let's take a moment here to demonstrate how problems where we are looking for fields within charge distributions can also be solved using the local form of Gauss's law. Using this method to solve for fields in empty space is fraught with mathematical nuance that we will avoid, but for regions containing charge it is quite workable, and perhaps even preferable in some cases.

Returning to the uniform sphere of charge, the spherical symmetry suggests that we write the divergence of the spherically-symmetric field in spherical coordinates. For vector fields that are only functions of r we have:

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}(r)=\frac{1}{r^2}\frac{d}{dr}\left[r^2E(r)\right] \tag{1.7.10}$$

We now apply Gauss's law and integrate. Note that this is an indefinite integral, which requires the introduction of an unknown constant of integration. To solve for this constant, we will need to know the *boundary condition* for the charge distribution. This is a universal feature of this method.

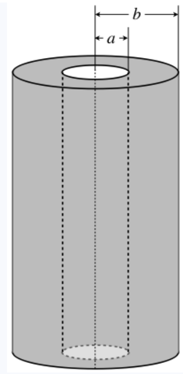
$$\frac{1}{r^2}\frac{d}{dr}\left[r^2E\right]=\frac{\rho}{\epsilon_o}\quad\Rightarrow\quad r^2E=\frac{\rho}{\epsilon_o}\int r^2dr=\frac{\rho r^3}{3\epsilon_o}+\beta\quad\Rightarrow\quad E(r)=\frac{\rho}{3\epsilon_o}r+\frac{\beta}{r^2} \tag{1.7.11}$$

Using the solution for outside the charge that we found above, and plugging in $r=R$ (the boundary), we find that our constant of integration comes out to be zero in this case. Note that we end up with the same field that we found using the integral version.

One thing to note about these two methods is that when the density is not constant, an integral has to be performed either way. Either the charge density appears in the integration of the divergence, or it appears in an integral to compute the charge enclosed within the volume enclosing (note that in this particular case of constant density we only had to multiply the density by the volume, but we will not always be so lucky).

Example 1.7.1

A very long insulating cylinder is hollow with an inner radius of a and an outer radius of b. Within the insulating material the volume charge density is given by: $\rho(R)=\alpha/R$, where α is a positive constant and R is the distance from the axis of the cylinder. Choose appropriate gaussian surfaces and use Gauss's law to find the electric field (magnitude and direction) everywhere.



Solution

There are three distinct regions: $(0 < r < a)$, $(a < r < b)$, *and* $(b < r < \infty)$. *For all of these regions, the radial symmetry of the charge distribution ensures that wherever there is electric field, it must point radially outward or inward, and its magnitude must be the same at every point at any fixed radius. A gaussian surface in the shape of a cylinder of radius* r *centered within the empty center region would therefore result in a flux of EA, where A is the area of the curved part of this surface, since the electric field is parallel to the area vector everywhere and constant in magnitude. We take each region in turn:*

*** $0 < r < a$ ***

The enclosed charge is zero, and since the area isn't zero, the electric field must be zero for every r *in that empty region.*

*** $a < r < b$ ***

The gaussian surface has a radius r *and a length* l. *The total electric flux is therefore:*

$$\Phi_E = EA = 2\pi r l E$$

To apply Gauss's law, we need the total charge enclosed by the surface. We have the density function, so we need to integrate it over the volume within the gaussian surface to get the charge enclosed. We use a volume in cylindrical coordinates ($dV = RdR\, d\theta\, dz$), *and the limits of integration are:* $R: a \rightarrow r$, $\theta: 0 \rightarrow 2\pi$, $z: 0 \rightarrow l$:

$$Q_{encl} = \int \rho dV = \int_0^l dz \int_0^{2\pi} d\theta \int_a^r \frac{\alpha}{R} RdR = 2\pi\alpha l\left(r-a\right)$$

Applying Gauss's law gives:

$$E\left(r\right) = \frac{Q_{encl}}{\epsilon_o A} = \frac{2\pi\alpha l\left(r-a\right)}{\epsilon_o 2\pi r l} = \boxed{\frac{\alpha}{\epsilon_o}\left(1-\frac{a}{r}\right)}$$

*** $b < r < \infty$ ***

With the gaussian surface now outside the entire charge distribution, the enclosed charge distribution is all of the charge. We can re-use the work above by simply changing the upper limit in the integral for enclosed charge from r *to* b. *This gives:*

$$Q_{encl} = 2\pi\alpha l\left(b-a\right) \quad \Rightarrow \quad E\left(r\right) = \boxed{\frac{\alpha}{\epsilon_o}\left(\frac{b-a}{r}\right)}$$

Example 1.7.2

Repeat the previous example for the outer two regions using the local form of Gauss's law. You can assume that you have already determined that $E = 0$ *in the hollow cavity, and use this as a boundary condition.*

Solution

In cylindrical coordinates, the divergence of a vector field that is only a function of the distance from the z-axis is given by:

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}(r)=\frac{1}{r}\frac{d}{dr}\left[rE(r)\right]$$

Now for each of the two regions we apply Gauss's law:

*** $a<r<b$ ***

We are given the function of the charge density in this region, so plugging that into the divergence formula gives:

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}(r)=\frac{\rho}{\epsilon_o}\quad\Rightarrow\quad\frac{1}{r}\frac{d}{dr}\left[rE(r)\right]=\frac{\alpha}{\epsilon_o r}\quad\Rightarrow\quad\frac{d}{dr}\left[rE(r)\right]=\frac{\alpha}{\epsilon_o}$$

Now perform the indefinite integral (don't forget the constant of integration! – I will call it β):

$$rE(r)=\int\frac{\alpha}{\epsilon_o}dr=\frac{\alpha}{\epsilon_o}r+\beta\quad\Rightarrow\quad E(r)=\frac{\alpha}{\epsilon_o}+\frac{\beta}{r}$$

The boundary condition at $r=a$ requires that the electric field is continuous there, which means that it must equal zero there. This allows us to solve for the constant of integration:

$$E(a)=0=\frac{\alpha}{\epsilon_o}+\frac{\beta}{a}\quad\Rightarrow\quad\beta=-\frac{\alpha a}{\epsilon_o}$$

Plugging this back in gives us the electric field in the region of the insulator, which agrees with the answer from the previous example:

$$E(r)=\frac{\alpha}{\epsilon_o}\left(1-\frac{a}{r}\right)$$

*** $b<r<\infty$ ***

There is no charge in this region, so the charge density is zero. Plugging this into the divergence formula gives:

$$\overrightarrow{\nabla}\cdot\overrightarrow{E}(r)=0\quad\Rightarrow\quad\frac{d}{dr}\left[rE(r)\right]=0\quad\Rightarrow\quad rE(r)=constant=\beta\quad\Rightarrow\quad E(r)=\frac{\beta}{r}$$

Once again we need to apply a boundary condition to determine β. In this case, we match the solution outside the cylinder to that inside the insulator region at $r=b$:

$$E(b)=\frac{\alpha}{\epsilon_o}\left(1-\frac{a}{b}\right)=\frac{\beta}{b}\quad\Rightarrow\quad\beta=\frac{\alpha}{\epsilon_o}(b-a)\quad\Rightarrow\quad E(r)=\frac{\alpha}{\epsilon_o}\left(\frac{b-a}{r}\right)$$

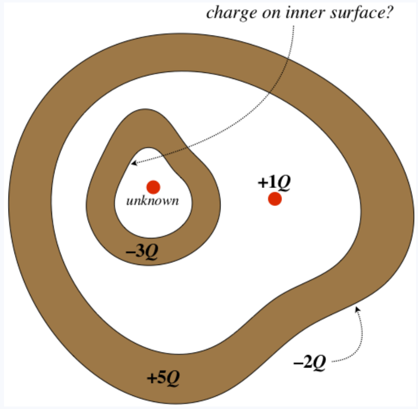
This again agrees with the answer obtained above.

Hollow Conducting Shells

Another common type of problem one can solve with Gauss's law involves no symmetry. A hollow conducting shell (of any shape) has an interesting property: Since the electric field within the metal of such a shell must vanish, then constructing a gaussian surface within the metal, we can see that there must be zero net flux through that surface. This means that the charge enclosed by that shell must be zero. But what if an electric charge is placed in the hollow space? How can Gauss's law then be satisfied? The only way is for charge in the conductor to migrate. If there is a positive charge in the hollow space, then an equal amount negative charge moves to the *inside surface* of the conducing shell, bringing the total charge within the gaussian surface to zero. If the shell was originally neutrally-charged, then the positive charge abandoned by the migrating negative charge resides on the outer surface of the shell.

Example 1.7.3

A closed, hollow conductor contains a smaller, closed hollow conductor and a point charge of $+1Q$ *(see the diagram). Free charge trapped within the interior space of the smaller conductor is unknown, but the smaller conductor itself carries a net charge of* $-3Q$*, and the larger conductor carries a net charge of* $+5Q$*. If a charge of* $-2Q$ *is found on the outside surface of the larger conductor, find how much charge resides on the inside surface of the smaller conductor. Assume all the charges on the conductors are at rest at equilibrium.*



Solution

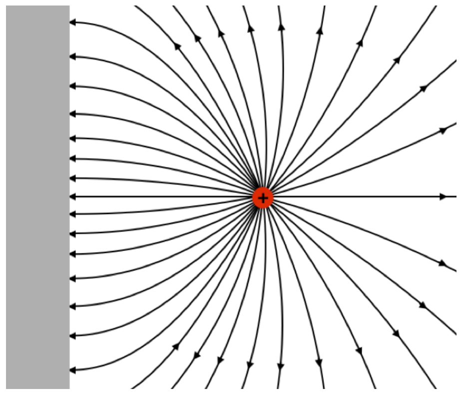
The total charge on the outer conductor must reside on its surfaces, so if $-2Q$ *is on the outer surface, then there must be* $+7Q$ *on its inner surface. Now construct a gaussian surface within the metal of the outer conductor. The zero electric field within the conductor (the charges are static) results in zero flux out of this gaussian surface, which means that there must be no net charge enclosed. The enclosed charge comes in many pieces, and is the sum of the charge on the inner surface of the outside conductor* ($+7Q$)*, the free charge outside the smaller conductor* ($+1Q$)*, the total charge on the smaller conductor* ($-3Q$)*, and the unknown free charge within the smaller conductor. For all of these to add up to zero, the unknown charge must be* $-5Q$*. There is also no flux through the inner conductor, so the charge enclosed within gaussian surface constructed within its metal must also be zero. Now that we know the previously-unknown charge is* $-5Q$*, there must be a charge of* $+5Q$ *on the inner surface of the smaller conductor.*

1.8: Method of Images

A Familiar-Looking Field

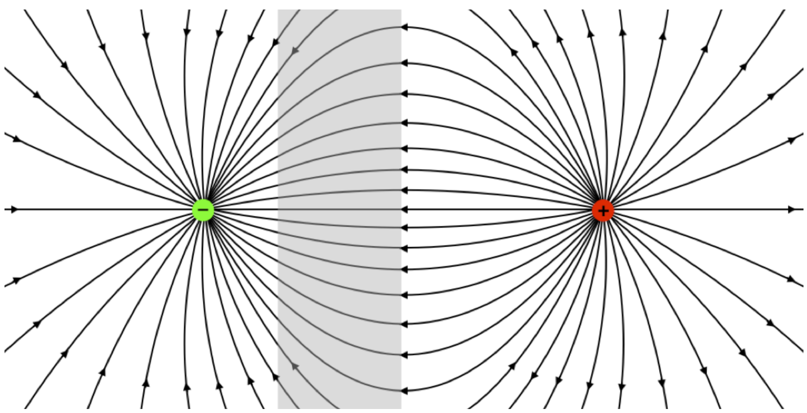
We are interested in seeing how the presence of a conductor affects the electric field, so we begin with the simplest case – a point charge near an infinite conducting plane. The key property of the conductor is that the field must strike it at right angles, which means that with a nearby point charge the field must look like:

Figure 1.8.1 – Field of a Point Charge Near a Flat Conductor



This field actually looks very much like (half of) a field that we have seen before:

Figure 1.8.2 – Field of Point Charge + Flat Conductor = Half a Dipole Field



It turns out that this is more than a resemblance of fields. In the region outside the conductor, the field is *precisely* the same as if the conductor were removed and a second point charge of equal magnitude and opposite sign was symmetrically-placed on the opposite side of where the conducting surface was.

A Clever Trick

The equivalence of these two fields provides us with an opportunity to use a clever trick for analyzing physical situations involving electric charges near flat conductors. For a point charge, this trick involves introducing an imaginary *image charge* reflected across the conducting surface, and using that charge to derive the actual field outside the conductor surface.

> **Alert**
>
> *It can't be stressed enough that this trick does not involve introducing an actual physical charge, any more than constructing a gaussian surface involved constructing an actual physical surface. These are techniques for performing calculations, and one should always keep in mind what the actual physical circumstances are.*

Now one might ask, "What's the big deal? All this trick gets us is the field strength for this single situation?" No, we get much more than this. First, consider that the charge outside the conductor will feel a force thanks to this field, *and we can actually calculate what this force is*, because it is the same as if the charge was in the presence of another point charge. If the charge is a perpendicular distance of x from the conductor, then the force the conductor exerts on the point charge is:

$$F = \frac{q^2}{4\pi\epsilon_o(2x)^2} = \frac{q^2}{16\pi\epsilon_o x^2} \tag{1.8.1}$$

Note that the force is attractive, and it would be attractive if the charge was negative as well, because the image charge is always the opposite sign of the actual charge. Also note that the force on this charge is exerted by the conductor (not by the image charge, which isn't actually there), so by Newton's third law, the force on the conductor is equal and opposite to the force on the point charge.

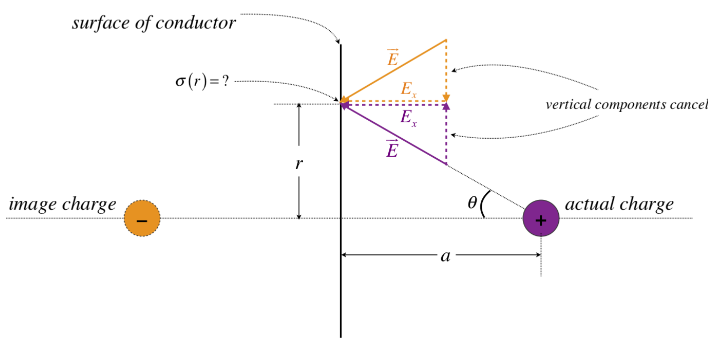
The power of this trick goes beyond what we can say about a single point charge. As we saw when we integrated fields for charge distributions, we can use superposition to determine a field of many point charges. If several point charges (or a continuous distribution of charge) is located near a flat conductor, then simply introducing an image charge for every charge (whether it is an individual point charge or an infinitesimal chunk of charge from a distribution) will work as a substitute for the conductor.

Charge on the Conductor

Let's return to the case of a point charge near the conducting plane. We know it is attracted to the conductor, but if the image charge isn't really there, what exactly is exerting this force? The answer must be charge, of course, but the conductor is neutrally-charged. But this charge is *free to move*, so naturally charge of the opposite sign as the point charge will migrate to the surface closest to the point charge, while charge of the opposite sign migrates the opposite direction. The pull of the closer charge is then greater than the push of the more distant charge, causing a net attractive force.

We can ask what the distribution of the charge on the conductor is, and one might be tempted to think that it is evenly-distributed, but this is not the case. We know that the charge density on the surface of a conductor is related to the field there by Equation 1.7.4, and we can use the image charge to determine the field at all positions on the conductor. Clearly the strength of the electric field at the surface has a circular symmetry, meaning that if we draw a line perpendicular to the conducting plane and through the charge, then the field at the surface is the same at all points equidistant from that line. Using the field of the real and image charge and Equation 1.7.4, we can therefore compute the charge density on the conductor as a function of r, the distance from the perpendicular line through the charge. Defining the distance separating the conductor and the charge to be a, we get the following diagram:

Figure 1.8.3 – Computing the Charge Density on the Conductor



The net electric field at the surface of the conductor is a sum of the x-components of the fields of the real and image charges, while the y-components of those fields cancel. The electric field magnitude for each charge comes from the coulomb field. Putting this all together gives:

$$E = 2E_x = 2E\cos\theta = 2\left[\frac{Q}{4\pi\epsilon_o\left(r^2+a^2\right)}\right]\left[\frac{a}{\sqrt{r^2+a^2}}\right] \quad \Rightarrow \quad \sigma\left(r\right) = \epsilon_o E\left(r\right) = \frac{-Qa}{2\pi\left(r^2+a^2\right)^{\frac{3}{2}}} \tag{1.8.2}$$

The minus sign was added to account for the fact that the sign of the charge on the surface is opposite the sign of the point charge Q. We can even determine *how much total charge* is brought to the near surface of the conductor. We get this by integrating the surface charge over the whole surface:

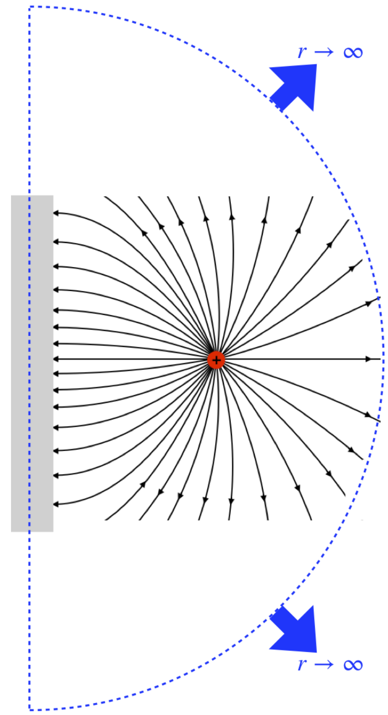
$$charge\ on\ surface = \int_{\substack{whole\\surface}} \sigma dA \tag{1.8.3}$$

This has radial symmetry, so we use polar coordinates. An element of area is $dA = rdrd\theta$, so:

$$charge\ on\ surface = \int_0^{2\pi} d\theta \int_0^{\infty} \sigma\left(r\right) rdr = -Qa\int_0^{\infty} \frac{r}{\left(r^2+a^2\right)^{\frac{3}{2}}} dr = -Qa\left[-\left(r^2+a^2\right)^{-\frac{1}{2}}\right]_0^{\infty} = -Qa\left[\frac{1}{a}\right] = -Q \tag{1.8.4}$$

So the amount of charge attracted to the near surface is exactly the negative of the point charge! There is actually a very clever alternative means of showing this that doesn't involve an integral, using Gauss's law. Construct a half-spherical gaussian surface of infinite radius with the flat part within the metal of the conducting plane, as depicted below.

Figure 1.8.4 – Gaussian Surface for Point Charge and Conducting Plane



The flat part of this gaussian surface permits zero flux, since the electric field inside the conductor vanishes, but what about the curved part of the gaussian surface? Well, the angles between the electric field lines and the surface vary wildly from one point on the hemisphere to another, but what about the *magnitude* of the electric field? This field is precisely the same as that of a dipole, which we already know from Equation 1.4.10 weakens with an inverse-cube law far from the source:

$$E_{large\ r} \propto \frac{1}{r^3} \tag{1.8.5}$$

The area of this hemisphere grows with the square of the radius:

$$A_{large\ r} \propto r^2 \tag{1.8.6}$$

The flux through any little part of the surface is *no larger* than the product of the magnitude of the electric field and the area (it is smaller when the angle the field makes with the area vector is greater than zero), so total flux changes with r no faster than:

$$\Phi_E\left(curved\ surface\right) \le EA \propto \frac{r^2}{r^3} = \frac{1}{r} \tag{1.8.7}$$

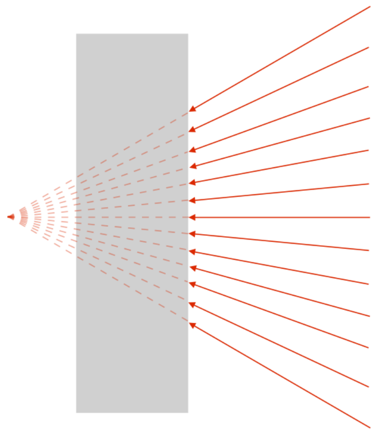
But as we are letting $r \to \infty$, this flux goes to zero, which when added to the zero flux through the flat surface, means that the total flux though this surface is zero. Gauss's law therefore insists that the enclosed charge is zero. The point charge contributes Q to the interior, which means that the conducting surface (also enclosed by the gaussian surface) must contribute $-Q$.

Field on the Other Side of the Conductor

We know that the image charge is a fabrication of our imaginations, but what actually is happening on the other side of this conductor? There is clearly Q on the far side, left behind by the $-Q$ that migrated to the near side (the conductor is neutrally-charged overall), so what kind of field does this create? Let's start by confirming that the field inside the conductor vanishes as it should.

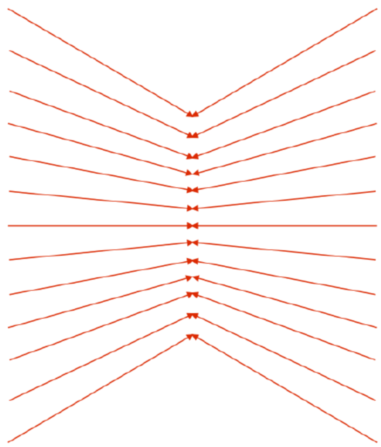
Begin by considering what the field of *just the charges on the near surface of the conductor* looks like outside that surface. We know that these charges create a field that (in that region) looks identical to the field created by the image charge, so it looks like:

Figure 1.8.5a – Field of the Surface Charge Alone



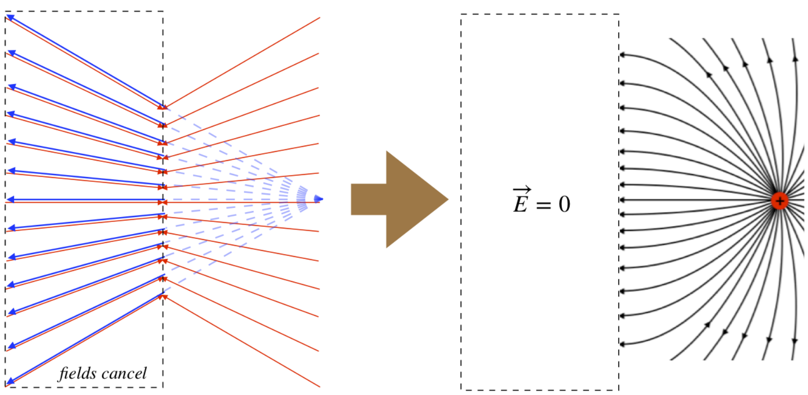
Next imagine removing the conductor, but leaving the surface charge fixed in place. Then the field would look the same on both sides of the charge:

Figure 1.8.5b – Field of the Surface Charge Alone



And now into this field, let's place our real positive charge in its proper position. Since the field lines on the right side of the figure above converge to the position of the image charge, by symmetry the field lines on the left side must converge to the position of the real charge. This means that the field on the left side above cancels the real point charge field in that region, giving a total field of zero to the left, and (as we have already shown) the half-dipole field to the right:

Figure 1.8.6 – Superposition of Real Charge's Field with Surface Charge's Field



So the charges are distributed on the surface of the conductor such that their field cancels the field of the point charge everywhere to the left of the surface. This confirms one thing we already know, which is that the field within the metal is zero. But it also proves that the charge left on the other side of the conductor is not affected at all by the fields of the point charge or the other surface charge. These orphaned charges will only be affected by one another, and will therefore repel each other the best they can. this leads to a uniform distribution on the other side of the conducting plane. When the charge density is uniform, it is simply the total charge divided by the total area. But the total charge is finite (it equals Q), while the area is infinite, so the surface charge density on the other side of the conductor is zero, and since the field at the surface is proportional to the surface charge density, the electric field is also zero! It is interesting that there is the same amount of charge on both sides of the conductor, and it results in a field on one side, and no field on the other, where it is simply spread too thin.

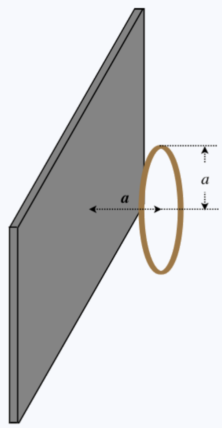
Applications

It should be apparent from the discussion above that the trick of using image charges (known as *the method of images*) has many applications for determining fields, forces, and surface charge densities. Specifically, we do not have to limit ourselves to the case of a single point charge. If there are multiple point charges present, we simply have to construct an image charge for each of them and use superposition. This can then be extended to continuous collections of charge as well – just construct an image collection of charge and use superposition from there.

For this class, we will limit ourselves to the simplest geometry of a single plane conductor, but the method can be expanded to multiple conducting planes and even spherical conductors.

Example 1.8.1

A thin circular plastic ring carries a net charge that is uniformly distributed throughout the ring with a linear density of λ. This ring is positioned parallel to a neutrally-charged infinite conducting plane such that its distance from the plane equals the radius of the ring.



In Example 1.3.1, we showed that the magnitude of the electric field on the axis of such a ring is given by:

$$E(x)=\frac{\lambda a x}{2\epsilon_o\left(a^2+x^2\right)^{\frac{3}{2}}},$$

where x is the distance from the ring along the axis. The field points along the axis of the ring, toward the ring (if the charge is negative), or away from it (if the charge is positive). A charged point particle is placed at the center of the ring. This particle has the same total charge as the ring, but it has the opposite sign. Find the magnitude and direction of the net force on this particle in terms of λ.

Solution

We can use the method of images to replace the conductor with image charges placed symmetrically on the opposite side of the conducting surface. We need images for both the point charge and the ring. The plastic ring will not contribute to the force on the point particle, because the particle is at its center, where the electric field of the ring vanishes. So the only two contributors to the force are the image point charge and the image ring. Start by determining the amount of charge on the particle in terms of the given linear density:

$$Q=\lambda l=2\pi a\ \lambda$$

The force between the point particle and its image is just the Coulomb force (with a separation of $r=2a$), and it is attractive (toward the plane), since the image charge has the opposite sign:

$$F_1=\frac{Q^2}{4\pi\epsilon_o(2a)^2}=\frac{\pi\lambda^2}{4\epsilon_o}$$

The force on the point charge by the image ring is the product of Q and the field of the image ring, and it is repulsive (the charge in the image ring has the opposite sign of the real ring, which is the same sign as the point charge). The image ring is *a distance $x = 2a$ from the point charge, so:*

$$F_2 = QE_{image\ ring} = (2\pi a\ \lambda)\frac{\lambda a\,(2a)}{2\epsilon_o\left(a^2+(2a)^2\right)^{\frac{3}{2}}} = \frac{2\pi\lambda^2}{5\sqrt{5}\epsilon_o}$$

These forces are in opposite directions, so the net force on the point charge is:

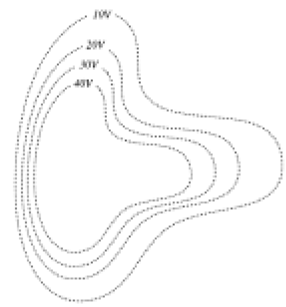
$$F = F_1 - F_2 = \boxed{\left(\frac{1}{4}-\frac{2}{5\sqrt{5}}\right)\frac{\pi\lambda^2}{\epsilon_o}}$$

The fact that this number is positive means that $F_1 > F_2$, so the net force is toward the conductor.

CHAPTER OVERVIEW

2: ELECTROSTATIC ENERGY

Now that we have described electrostatic phenomena in terms of vector fields that relate to the forces on charges, we now turn to scalar fields, which relate to the potential energies of charges.



2.1: POTENTIAL ENERGY OF CHARGE ASSEMBLY

Charges interact through electric forces, which do work when the charges are moved. Assembling collections of charge therefore results in a potential energy change.

2.2: ELECTROSTATIC POTENTIAL

We defined an electric vector field as the force on a charge divided by that charge, so that it depends only on the source charges. We now do the same to define a scalar potential field by dividing the potential energy of a charge by that charge.

2.3: COMPUTING POTENTIAL FIELDS FOR KNOWN CHARGE DISTRIBUTIONS

Just as we did for electric fields, we can calculate the potential field of a given charge distribution. While the procedures of these two calculations are similar, there are some important differences.

2.4: CAPACITANCE

Electrical potential energy is typically stored by separating oppositely-charged particles and storing them on different conductors. Such systems of energy-storing, oppositely-charged conductors are called capacitors.

2.5: DIELECTRICS

We defined a perfect insulator as a substance that doesn't allow for any movement of electric charge. But in fact while insulators don't allow charge to migrate freely, they do allow charges to displace slightly, and this affects the electric field within the substance.

2.6: STATIC NETWORKS

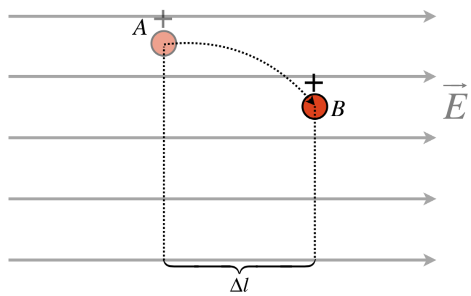
We move now into more practical considerations for capacitors, namely what happens when we actually connect them to each other and to batteries with conductors.

2.1: Potential Energy of Charge Assembly

Potential Energy of a Point Charge in a Field

In our brief discussion of the potential energy of dipoles in external fields in Section 1.4, we noted that an electric charge that is displaced within an electric field can have work done on it by the electric force, and this can be expressed as the negative of a change in electrical potential energy. Pruning-down Figure 1.4.5 to a single electric charge, we have:

Figure 2.1.1 – Change of Potential Energy for a Charge Displaced Within a Field



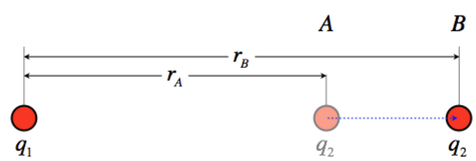
$$W_{A\rightarrow B}=\int_A^B\vec{F}\cdot\vec{dl}=\int_A^B\left(qE\;\hat{i}\right)\cdot\left(dx\;\hat{i}+dy\;\hat{j}\right)=\int_A^B qEdx=qE\Delta l\quad\Rightarrow\quad\Delta U=U_B-U_A=-W_{A\rightarrow B}=-qE\Delta l \tag{2.1.1}$$

This was easy enough to compute, since the electric field was uniform. We now look at cases where this is not the case. The simplest such case is changing the separation of two point charges. And since starting with point charges is always the basis for more general cases, this is the perfect place to start.

Potential Energy of a Multiple Point Charges

Consider two point charges, q_1 and q_2 that are separated by an initial distance r_A and are moved to a new separation r_B (we will assume that only q_2 moves). We seek the work done on q_2 during this move by the electric field coming from q_1, from which we can obtain the change in the system's potential energy.

Figure 2.1.1 – Change of Potential Energy for a Two Point Charges



The force between these charges changes as q_2 is moved, which means that the work calculation requires a far less trivial integral than was performed for the case of a uniform field. Start by setting up the work integral with the coluomb force:

$$\vec{F}_{on\;q_2}=\frac{q_1q_2}{4\pi\epsilon_or^2}\hat{r}\quad\Rightarrow\quad W_{A\rightarrow B}=\int_A^B\vec{F}\cdot\vec{dl}=\int_A^B\left(\frac{q_1q_2}{4\pi\epsilon_or^2}\hat{r}\right)\cdot\vec{dl} \tag{2.1.2}$$

The displacement is parallel to the radial unit vector (if it wasn't, the dot product would require that we only take the displacement in that direction anyway), so the product $\hat{r}\cdot\vec{dl}$ can be written simply as $+dr$. Putting this into the integral gives:

$$W_{A\rightarrow B}=\frac{q_1q_2}{4\pi\epsilon_o}\int_{r_A}^{r_B}\frac{1}{r^2}dr=\frac{q_1q_2}{4\pi\epsilon_o}\left[\frac{1}{r_A}-\frac{1}{r_B}\right] \tag{2.1.3}$$

The change in potential energy for this process is therefore:

$$\Delta U = -W_{A\to B} = \frac{q_1 q_2}{4\pi\epsilon_o}\left[\frac{1}{r_B} - \frac{1}{r_A}\right] \tag{2.1.4}$$

As we recall from our study of mechanics, it is only the change in potential energy that matters, but we also find it useful to define a state of zero potential, from which we can reference other states. It is common (though not universal, as well will see later) to reference our point of zero electrostatic potential energy at $r=\infty$. Another way to look at this is to think of the potential energy of a configuration of charges (in this case, two point charges) as the work done in moving the charges from infinite separation to their current proximity. That gives us the following potential energy of two point charges separated by a distance r:

$$U(r) = -W_{\infty\to r} = \frac{q_1 q_2}{4\pi\epsilon_o r} \tag{2.1.5}$$

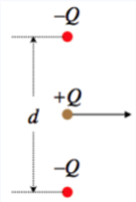
It should be noted that this potential energy is positive if the two charges have the same sign, and negative if they have different signs. This makes sense, since we have to add external work to the system to push the repelling charges together, while attracting charges "want" to come together, which is a characteristic of decreasing potential energy (because the force causes them to speed up, so the loss of potential energy results in a increase of kinetic energy).

When we collect more than just two point charges, we have to account for the potential energy contribution of every pair of charges. This begins to add up when the number of point charges grows. Representing the separation of charge 1 from charge 2 with "r_{12}", charge 1 from charge 3 with "r_{13}," and so on, the total potential energy for a collection of point charges is the sum of all the pairwise contributions:

$$U_{total} = \frac{q_1 q_2}{4\pi\epsilon_o r_{12}} + \frac{q_1 q_3}{4\pi\epsilon_o r_{13}} + \frac{q_2 q_3}{4\pi\epsilon_o r_{23}}\ldots \tag{2.1.6}$$

Example 2.1.1

A point charge Q is moving horizontally halfway between two other point charges that are equal in magnitude but opposite in sign, and are held fixed in place. The two negative point charges are separated by a distance d. The positive point charge obviously experiences no net vertical force, so it continues moving horizontally. Find the amount of KE gained or lost (indicate which) by the moving charge at the moment when the three charges form an equilateral triangle.



Solution

The mechanical energy will be conserved, so the change in kinetic energy will equal the negative the change of potential energy. The potential energy in the system that results from the two negative charges interacting with each other never changes (they are held in place), and the potential energy change between the moving charge each of the stationary charges is the same due to symmetry. So all we need to calculate is the change in potential energy between the moving charge and one of the others, and multiply it by two. They are initially separated by a distance $\frac{d}{2}$, and afterward are separated by d, so:

$$\Delta KE = -\Delta U_{system} = -2\left[\frac{-Q^2}{4\pi\epsilon_o d} - \frac{-Q^2}{4\pi\epsilon_o \frac{d}{2}}\right] = -\frac{Q^2}{2\pi\epsilon_o d}$$

The change in kinetic energy is negative, indicating that the charge slows down. This makes sense, given that it is attracted by the other charges, which pull it in the direction opposite to its motion.

Potential Energy of Continuous Distributions

We are of course not only interested in collections of finite numbers of point charges, but continuous distributions of charge as well. To compute the potential energy stored in collecting a distribution of charge can be tricky business. The idea is to set up an integral of contributions to the potential energy due to the addition of an infinitesimal charge.

The simplest example is the case of a conducting sphere of radius R and a surface charge Q (which, thanks to the sphere being a conductor, is evenly-distributed). To determine the potential energy stored in this system, we consider the incremental energy added to it when the surface of the sphere has some intermediate amount of charge q, and we bring dq from infinity to the surface. The field of the spherical shell of charge is identical to that of an equal amount of charge located at its center (which is easily shown using Gauss's law), so the potential energy change of bringing dq from infinity to its surface is:

$$dU = \frac{q dq}{4\pi\epsilon_o R} \tag{2.1.7}$$

Now we keep collecting charge like this until the total charge equals Q. We started with zero charge on the surface, so the limits of integration are 0 to Q:

$$U_{spherical\ shell} = \int_0^Q \frac{q dq}{4\pi\epsilon_o R} = \frac{Q^2}{8\pi\epsilon_o R} \tag{2.1.8}$$

Naturally the potential energy is positive regardless of the sign of Q, because work needs to be done to push together charges of the same sign, regardless of whether they are positive or negative.

We can do the same for a sphere that is uniformly filled with charge, though the procedure requires a bit more thought. Using the same total charge and radius as above, we begin by noting that the charge density within the sphere:

$$\rho = \frac{Q}{V} = \frac{Q}{\frac{4}{3}\pi R^3} = \frac{3Q}{4\pi R^3} \tag{2.1.9}$$

Now imagine building the solid sphere from the inside-out, one infinitesimally-thin shell at a time. All the charge on such a shell is the same distance from the center, and sees whatever charge is already present as if it was a point charge at the center. Calling the amount of charge already present q, the gain in potential energy that comes from adding the shell (which contains an infinitesimal amount of charge we'll call dq) is:

$$dU = \frac{q dq}{4\pi\epsilon_o r} \tag{2.1.10}$$

Notice that the distance to the center is r, not R. This is because the shells that are added are not yet out to the full sphere's radius. To integrate this, we need everything written in terms of a single variable, and the simplest to use is r, which will vary from 0 to R. The amount of charge within a sphere of radius r is:

$$q = \rho V = \rho\left(\frac{4}{3}\pi r^3\right) = Q\frac{r^3}{R^3} \tag{2.1.11}$$

The amount of infinitesimal charge in a spherical shell is the volume of that shell times the density. The volume of the shell is the surface area times the differential radius, so:

$$dq = \rho dV = \rho\left(4\pi r^2 dr\right) = 3Q\frac{r^2}{R^3}dr \tag{2.1.12}$$

Okay, putting all this together and integrating gives us our answer:

$$U_{uniform\ solid\ sphere} = \int_{r=0}^{r=R} \frac{q dq}{4\pi\epsilon_o r} = \int_{r=0}^{r=R} \frac{\left(Q\frac{r^3}{R^3}\right)\left(3Q\frac{r^2}{R^3}dr\right)}{4\pi\epsilon_o r} = \frac{3Q^2}{4\pi\epsilon_o R^6}\int_{r=0}^{r=R} r^4 dr = \frac{3Q^2}{20\pi\epsilon_o R} \tag{2.1.13}$$

Notice that despite having the same amount of charge and the same radius, there is more energy stored in this system than in that of a hollow shell. This makes sense if one imagines taking some of the charge on the surface of the hollow sphere and pushing it into the middle to make the sphere a continuous solid collection of uniform charge. Pushing the same-sign charges closer together involves doing work on the system, which adds potential energy to it.

Example 2.1.2

An insulating sphere of radius R contains a net charge that is non-uniformly-distributed. The charge is distributed in a spherically-symmetric manner, depending only upon the distance r from the center of the sphere, according to:

$$\rho(r) = \rho_o\,\frac{r}{R}$$

Find the energy stored in this configuration, in terms of the total charge Q present, and the radius of the sphere.

Solution

We start with a partially-assembled sphere with charge q, which occupies the sphere from the center to a radius r. This much charge can be written in terms of ρ_o and r:

$$q=\int\rho dV=\int_0^r\rho\left(r'\right)4\pi r'^2dr'=\frac{4\pi\rho_o}{R}\int_0^r r'^3dr'=\frac{\pi\rho_o}{R}r^4$$

We can write the constant ρ_o in terms of the total charge by integrating the entire sphere:

$$Q=\frac{\pi\rho_o}{R}R^4=\pi\rho_oR^3\quad\Rightarrow\quad\rho_o=\frac{Q}{\pi R^3}$$

So we can write the charge in the partially-filled sphere and the infinitesimal charge in the thin shell at the outer radius of the partial sphere in terms of the total charge:

$$q=Q\frac{r^4}{R^4}\;,\qquad dq=4Q\frac{r^3}{R^4}dr$$

Now we just have to integrate the potential energy function over the full assembly of charge:

$$U=\int_{r=0}^{r=R}\frac{qdq}{4\pi\epsilon_or}=\frac{Q^2}{\pi\epsilon_oR^8}\int_{r=0}^{r=R}r^6dr=\frac{Q^2}{7\pi\epsilon_oR}$$

This energy is higher than for the same amount of charge all on the surface, but lower than for the uniform distribution. This makes sense, since more of the charge has been pushed close together than the hollow shell, but the density gets smaller as we get closer to the center, so more charge is pushed together in the uniform case.

We have limited ourselves to the energy stored in the assembly of spherical charge distributions, thanks to the high degree of symmetry. But the process for less-symmetric assemblies works pretty much the same way, and we will soon see some additional tools that can help with this.

2.2: Electrostatic Potential

Test Charges

An alternative way to look at electric fields from what we did in Section 1.2 is from the perspective of a *test charge*. The idea is to use a charged point particle as a means of measuring electric force vectors at various points in space. When the force vectors are all mapped-out, we then divide them by the charge of the point particle, and the new vectors are then the electric field vectors. A common (but somewhat strange) way to write this mathematically is:

$$\vec{E}\left(\vec{r}\right)=\lim_{q_{test}\to 0}\frac{\vec{F}_{on\;q_{test}}}{q_{test}},\qquad \text{where }\vec{r}\text{ is the position of }q_{test}\tag{2.2.1}$$

A region around a collection of charge can similarly be tested with a charged point particle. At every point in space, the potential energy of the point charge can be measured, and then the amount of testing charge can be divided out, so that all that remains is a function of the source charges. We write it this way:

$$V\left(\vec{r}\right)=\lim_{q_{test}\to 0}\frac{U\left(q_{test}\right)}{q_{test}},\qquad \text{where }\vec{r}\text{ is the position vector of }q_{test}\tag{2.2.2}$$

This process maps out a *scalar field*, since at every point in space is associated a number (not a vector, like in the case of electric field). Just as electric field vectors are not the same as force vectors, the values in this scalar field are not potential energies – indeed, this can be seen even in the units of these numbers, which are joules divided by coulombs. The ratio of joules per coulomb is given its own name: *volts*. The scalar field we have invented this way is called *electrostatic potential*. Like an electric field vector, this is a quantity that is defined at every point in space in the vicinity of some electric charge. Unlike electric field vectors, these quantities are scalars – they have no direction.

> **Alert**
>
> *Possibly the most confusing thing to students new to electrostatics is use of the word "potential" in "electrostatic potential." This name derives from the fact that it is related to electric potential energy, but these quantities are very different, and the reader is advised to keep this in mind.*

Superposition

When there is more than one source of electric field in the vicinity of a point in space, the contributions of those sources to the field at that point can be added together. This can be seen simply from the test charge approach – clearly the forces on the test charge can be added together, and when the test charge is divided out, the sum of the electric field vectors remains. We see the same thing for electrostatic potential:

$$U\left(q_{test}\right)=\frac{q_1 q_{test}}{4\pi\epsilon_o r_1}+\frac{q_2 q_{test}}{4\pi\epsilon_o r_2}+\frac{q_3 q_{test}}{4\pi\epsilon_o r_3}\dots\quad\Rightarrow\quad V\left(\vec{r}\right)=\frac{U\left(q_{test}\right)}{q_{test}}=\frac{q_1}{4\pi\epsilon_o r_1}+\frac{q_2}{4\pi\epsilon_o r_2}+\frac{q_3}{4\pi\epsilon_o r_3}\dots\tag{2.2.3}$$

Here r_i is the distance from the i^{th} source charge to the position in space indicated by the position vector $\vec{r}$. Notice that by adopting the $U(\infty)=0$ convention, we have also done so for the electrostatic potential. And like the potential energy, the position that we choose to call the electric potential zero is arbitrary.

All of the things we developed for electric fields also apply to potentials, with the only difference being that potentials superpose as scalars, not vectors (which actually makes them easier to deal with in many cases). The main point is that when we have a collection of source charges – including a continuous distribution – we can define a potential at every point in space, and if we place a point charge there, we can determine its potential energy by multiplying the charge by the electric potential:

$$U=qV\left(\vec{r}\right),\qquad \text{where }\vec{r}=\text{position vector of the charge }q\tag{2.2.4}$$

The similarity with Equation 1.2.2 is obvious – we have simply replaced force and field with energy and potential.

> **Alert**

We will frequently use the language like, "the potential energy of the point charge," but as with all potential energy, we really mean, "the potential energy added to the system thanks to the presence of the point charge." It takes an interaction through a conservative force to introduce potential energy, and interactions require two entities. For example, an object cannot have its own gravitational potential energy (though we often treat it that way) – it needs to interact with the Earth.

Bridging the Scalar and Vector Fields

It's clear that both the scalar potential field and the vector electric field are determined by the source charge distribution, so these fields must be related to each other somehow. Indeed they are! To see how, we once again look back to our study of mechanics, where we related potential energy and force. We've already mentioned one such relation, through the computation of work:

$$\Delta U = U_B - U_A = -W_{A\rightarrow B} = -\int_A^B \overrightarrow{F}\cdot\overrightarrow{dl} \tag{2.2.5}$$

If we just plug in $U = qV$ and $\overrightarrow{F} = q\overrightarrow{E}$, we get a direct relation between the change of potential and the electric field:

$$V_A - V_B = \int_A^B \overrightarrow{E}\cdot\overrightarrow{dl} \tag{2.2.6}$$

[*It is actually common to use the units of Vm^{-1} for electric field, rather than our previous NC^{-1}.*]

Note that the signs have been flipped on both sides of the equation. The quantity on the left is usually referred to as the *potential drop from A to B*. Of course, the potential doesn't *have* to drop, so perhaps potential *change* is better language. The reason for this wording probably has its roots in the specific case of performing the integral along a path that follows the direction of the electric field. Notice that in this case, $\overrightarrow{E}$ is always in the same direction as $\overrightarrow{dl}$, which gives a positive line integral. If the line integral is positive, then $U_A > U_B$, which means that the potential drops from A to B. This gives us a useful rule of thumb:

Electric fields point in the direction of decreasing electric potential.

Whenever we have an integral relationship like this, then as we saw for Gauss's law, a differential (local) relation is also available. We actually saw this back in our study of mechanics, and it comes through here as well:

$$\overrightarrow{F} = -\overrightarrow{\nabla}U \quad \Rightarrow \quad \overrightarrow{E} = -\overrightarrow{\nabla}V \tag{2.2.7}$$

While this is interesting, the reader can be forgiven for asking what use it has. This last relation is particularly powerful for the following reason. Suppose we wish to compute the electric field of a charge distribution. Assuming we don't have a clever way of using Gauss's law to do this, we have to perform a calculation like we did back in Section 1.3. Part of what makes that computation challenging is keeping track of three different components of the electric field vector (i.e. three separate integrals). If we instead compute the potential field (one integral, with no vectors involved), we can then take *derivatives* (the gradient) to get the electric field. We will see how one calculates the potential field from a distribution of charge in the next section.

Alert

*The relation between field and potential is often misunderstood, in yet another incarnation of confusing a quantity with a change in that quantity (like mistaking acceleration with velocity. Just as zero instantaneous velocity does not mean the acceleration is zero, a zero potential at a point in space does not mean that the field there is zero. Indeed, we can define the potential to be zero anywhere, no matter what the field is! It is the **rate of change** of the potential that determines the field, not the **value** of the potential.*

Gradient Formulas

Back in Section 1.6 we encountered our first use of vector calculus when we learned that we would have to take divergences of electric fields to apply Gauss's law in certain applications. Now we are faced with one of the cousins of the divergence operation – the gradient. As we did with divergence, it is useful to review some formulas for gradients in certain special circumstances. As with the divergence, the formula for the gradient in cartesian coordinates works in all cases, while the gradient in cylindrical and spherical coordinates are only simplified when the scalar function depends only upon r (as before, in cylindrical coordinates, this is the distance to an axis, and in spherical coordinates it is the distance to a point):

Cartesian Coordinates

$$\overrightarrow{\nabla} V\left(x,\ y,\ z\right)=\frac{\partial V}{\partial x}\hat{i}+\frac{\partial V}{\partial x}\hat{j}+\frac{\partial V}{\partial x}\hat{k} \tag{2.2.8}$$

Cylindrical Coordinates

$$\overrightarrow{\nabla} V\left(r,\ \not{\phi},\ \not{z}\right)=\frac{\partial V}{\partial r}\hat{r} \tag{2.2.9}$$

Spherical Coordinates

$$\overrightarrow{\nabla} V\left(r,\ \not{\theta},\ \not{\phi}\right)=\frac{\partial V}{\partial r}\hat{r} \tag{2.2.10}$$

Example 2.2.1

In a certain region of space around the origin, the electrostatic potential field satisfies:

$$V\left(x,y,z\right)=\alpha\ x+\beta\ y^2+\gamma\ z^3$$

Find the charge density at the origin in terms of α, β, and γ.

Solution

We can find the electric field from the potential field:

$$\overrightarrow{E}=-\overrightarrow{\nabla}V=-\frac{\partial V}{\partial x}\ \hat{i}-\frac{\partial V}{\partial y}\ \hat{j}-\frac{\partial V}{\partial z}\ \hat{k}=-\alpha\ \hat{i}-2\beta\ y\ \hat{j}-3\gamma\ z^2\ \hat{k}$$

Now the divergence of the field gives us the charge density (Gauss's law in local form):

$$\frac{\rho\left(x,y,z\right)}{\epsilon_o}=\nabla\cdot\overrightarrow{E}=\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}+\frac{\partial E_z}{\partial z}=0-2\beta-6\gamma\ z\quad\Rightarrow\quad\boxed{\rho\left(0,0,0\right)=-2\beta\epsilon_o}$$

Notice that at the origin the potential is zero, but the electric field is not, nor is the charge density.

An Identity from Vector Calculus

We derived the negative-gradient relationship between the potential and the electric field from the same relation between a conservative force and its associated potential energy (Equation 2.2.7). For a force to have an associated potential energy, it is necessary that it be conservative. We have been assuming all along that the electric force is conservative. As we will see later, this is actually not always the case. It turns out that the *electromagnetic field* is conservative, but it is possible for the magnetic field to transfer energy to/from the electric field, making the electric field by itself not conservative. For this to be the case, the source charges need to be moving, and since we are still discussing only electro*statics*, we can safely continue to use the electrostatic potential and the negative gradient relation.

The existence of a potential energy function is sufficient to prove that a force is conservative, though proving this can be troublesome, without the tools provided by vector calculus. The simple test for whether a force is conservative is if its curl vanishes:

$$\overrightarrow{\nabla}\times\overrightarrow{F}=0\quad\leftrightarrow\quad F\ is\ conservative\quad\leftrightarrow\quad\overrightarrow{F}=-\overrightarrow{\nabla}U \tag{2.2.11}$$

The reason this works as a test is that the geometry of the curl and gradient are such that the curl of a vector field that comes from a gradient of a scalar field is always identically zero:

$$\overrightarrow{\nabla}\times\left[\overrightarrow{\nabla}\left(anything\right)\right]\equiv 0 \tag{2.2.12}$$

So if the force can be written as the negative gradient of a potential energy function, then its curl must vanish, and this corresponds to a conservative force. Extending this to electrostatics, we see that if the electric field can be expressed as the negative gradient of a potential, then its curl vanishes. So we have:

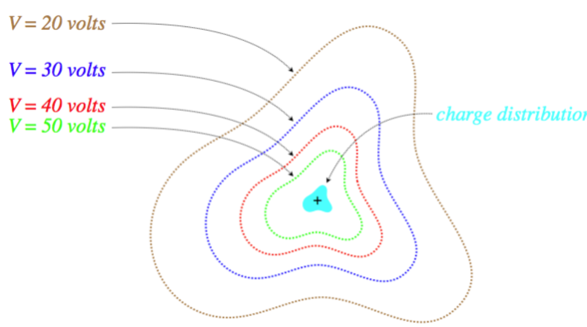
$$\vec{\nabla} \times \vec{E} = 0 \quad \leftrightarrow \quad electrostatics \quad \leftrightarrow \quad \vec{E} = -\vec{\nabla} V \tag{2.2.13}$$

Equipotential Surfaces

A consequence of the gradient relation is that their relationship is geometric in nature. The first manifestation of this is that the gradient of a scalar field points in the direction where the scalar values are increasing the fastest. With the presence of the negative sign, we therefore conclude that the electric field points in the direction of the fastest *descent* of electric potential. This confirms the rule-of-thumb we established above.

We can demonstrate this geometrical relationship through a diagram. Let's imagine starting at a certain point in space, and measuring the potential there (after designating the zero point). Then we sample nearby points, and find a direction we can move our detector so that the potential doesn't change. If we keep following this procedure, and map the entire space where the potential doesn't change, we will find that it is a surface. As this imaginary surface exists at a single, equal potential, it called an *equipotential surface*. Here is a two-dimensional depiction of a collection of such surfaces:

Figure 2.2.1 – Equipotential Surfaces



With a positive source charge, the field lines are pointing outward, which is indeed pointing from higher potential to lower potential, but there is something more specific that we can conclude about the geometric relationship of the field and potential. When we follow a path that remains on an equipotential, the potential never changes, so if we traverse such a path from position A to position B, we find:

$$V_A - V_B = 0 = \int_A^B \vec{E} \cdot \vec{dl} \tag{2.2.14}$$

Certainly the electric field is not zero everywhere we go, and the distance we travel isn't zero, so how can this integral come out to be zero? Maybe parts of it cancel other parts? No, because it happens on every single path we take, between any two points, so long as that path stays on an equipotential. The answer is that the only way this integral can be zero is if at every point on the equipotential, the electric field is perpendicular to $\vec{dl}$. In other words:

The electric field is perpendicular to equipotential surfaces everywhere.

That electric fields are perpendicular to equipotential surfaces sounds very familiar. We said the same thing about conducting surfaces for electrostatics. Indeed, we immediately conclude that for electrostatics:

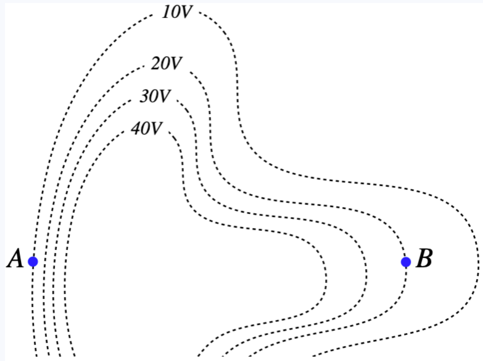
Conductors are equipotentials.

Note that this statement goes beyond just the surface of the conductor. We know that inside the metal of the conductor there is no electric field, so as we go from the surface of the conductor into the metal, the electric potential can't be changing (electric fields come from changes of electric potential), so the electric potential is the same everywhere in the conductor.

When we are provided with several equipotential surfaces as we are here, we can conclude more about the electric field than just its direction. The gradient operation measures a directional *rate of change*. This means that if every equipotential surface shown is separated by the same number of volts (as in the diagram above), then the regions where those surfaces are closest together is where they are changing the fastest, which means that the magnitude of the electric field is greatest there.

Example 2.2.2

A charged particle travels through an electric field whose equipotential surfaces are shown in the diagram. The only force experienced by the charge is due to this field. The charge is moving slower at point A than it is at point B.



a. *The charge of the particle is: positive / negative / can't tell*
b. *The magnitude of the charge's acceleration is greater at: point A / point B / can't tell.*
c. *What is the direction of the charge's acceleration at each point?*

Solution

a. The particle's kinetic energy increased from point A to point B, which means that its potential energy went down. But its electrostatic potential went up, so since $\Delta U = q\Delta V$, then $\Delta U < 0$ and $\Delta V > 0$ means that $q < 0$.

b. The equipotentials all differ by equal voltages, so those that are closer together indicate a region where the electric field is stronger. The field is therefore stronger at point A, which means it experiences a greater net force there than it does at point B.

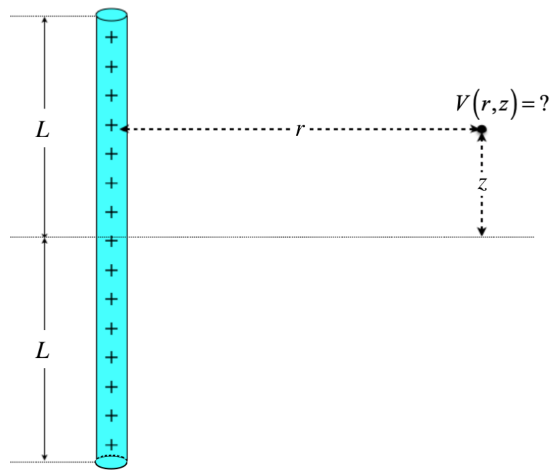
c. The force due to the electric field must be parallel to the electric field, which must be perpendicular to the equipotential surface. So the forces at points A and B must be either to the left or to the right, but can we tell which way? The field points from higher potential to lower potential, so at point A it points left, and at point B is points right. The charge is negative, so the forces are opposite to the electric field directions. The particle accelerates to the right at point A and to the left at point B.

2.3: Computing Potential Fields for Known Charge Distributions

Electrostatic Potential Field of Straight Line Segment of Uniform Density

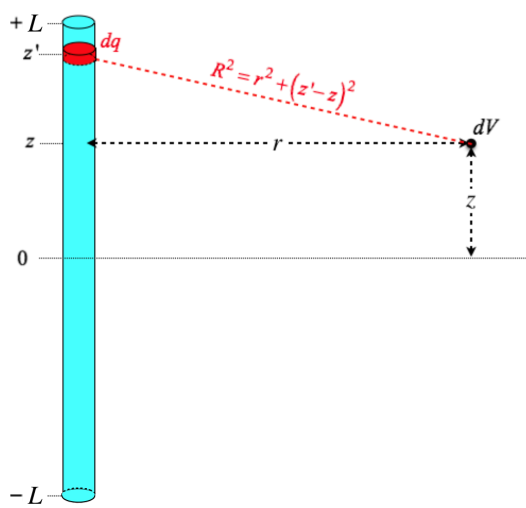
We return to the problem we first solved way back in Section 1.3. Actually, we are going to take it a step further than we did before, which will show the power of approaching the problem of computing fields with potentials. In our direct calculation of the electric field, we punted the computation of the z-component due to the lack of symmetry. It's not that it was impossible to do, but the vector element of the calculation was daunting, and in any case we now have a different approach that is easier to use. So the physical set-up is the same as before – a uniform linear segment of charge located along the z-axis from $z=-L$ to $z=+L$. But now we seek the potential field at a position a distance r from the z-axis, *and* a distance z from the (x,y) plane. We still have cylindrical symmetry, which means we don't have to specify the polar angle of the position of the point in space.

Figure 2.3.1a – Electrostatic Potential Field of a Uniform Line Segment



We start the same way as we did for the electric field – identifying an element of charge, setting up a coordinate system, and determining the distance from the point charge to the point in space:

Figure 2.3.1b – Electrostatic Potential Field of a Uniform Line Segment



To start the math, we make several notes: First, unlike the electric field, we are not dealing with vectors, so we don't have to track components (yay!). Second, we need to reference the zero point of potential, and we will do so with the usual $V(\infty)=0$. Finally, it should be noted that it is a bit awkward to integrate over the z variable, while our final answer is a function of the z component of the position in space, so we have named our integration variable z'.

$$\left.\begin{array}{l} dq=\lambda\; dz' \\ dV=\dfrac{dq}{4\pi\epsilon_o R} \end{array}\right\} \quad\Rightarrow\quad V(r,z)=\int\limits_{-L}^{+L}\frac{\lambda\; dz'}{4\pi\epsilon_o\sqrt{r^2+(z'-z)^2}} \tag{2.3.1}$$

With a substitution of $u \equiv z' - z$, the integral becomes easy to look up in an integral table:

$$V(r,z) = \frac{\lambda}{4\pi\epsilon_o} \int\limits_{-z-L}^{-z+L} \frac{du}{\sqrt{r^2+u^2}} = \frac{\lambda}{4\pi\epsilon_o} \left[\ln\left(\sqrt{r^2+u^2}+u\right)\right]_{-z-L}^{-z+L} = \frac{\lambda}{4\pi\epsilon_o} \ln\left(\frac{\sqrt{r^2+(z-L)^2}+L-z}{\sqrt{r^2+(z+L)^2}-L-z}\right) \tag{2.3.2}$$

Okay, so that's quite a mess, but consider this: The set-up and the math to get here was not all that tough, *and* we have so much more information here. This is the potential field *throughout all of space*. When we calculated the electric field for this charge distribution, we were somewhat daunted by the lack of symmetry that occurs off the x-axis, making dealing with the vector components off the x-axis problematic (not impossible, but certainly not much fun). And now if we want the electric field at any point in space (the $\hat{z}$ component as well as the $\hat{r}$ component!), we only need to take derivatives, since the potential field is related to the electric field through the gradient.

Example 2.3.1

If we take the solution found in Example 1.3.2 and make the cylinder infinitesimally thin ($a \to 0$), we get that the electric field magnitude on the z-axis for a uniformly charged rod of length $2L$ is:

$$E = \frac{Q}{8\pi\epsilon_o L}\left[\frac{1}{z-L} - \frac{1}{z+L}\right]$$

a. *Find the electrostatic potential on the z-axis for this collection of charge.*
b. *Use the electrostatic potential on the z-axis to show that the electric field there comes out to what we found previously.*

Solution

a. We can start from scratch an perform the integral, but why do this, when we have a general solution for this charge distribution above? All we need to do is take the limit of $r \to 0$ and we will be confined to the z-axis:

$$\lim_{r\to 0} V(r,z) = \frac{\lambda}{4\pi\epsilon_o} \lim_{r\to 0} \ln\left(\frac{\sqrt{r^2+(z-L)^2}+L-z}{\sqrt{r^2+(z+L)^2}-L-z}\right) = \frac{\lambda}{4\pi\epsilon_o} \ln\left(\frac{\sqrt{0+(z-L)^2}+L-z}{\sqrt{0+(z+L)^2}-L-z}\right) = \frac{\lambda}{4\pi\epsilon_o} \ln\left(\frac{0}{0}\right)$$

We have an indeterminate form, so we need to employ l'Hôpital's rule:

$$\lim_{r\to 0} \ln\left[\frac{f(r)}{g(r)}\right] = \ln\left[\lim_{r\to 0}\frac{f(r)}{g(r)}\right] = \ln\left[\lim_{r\to 0}\frac{f'(r)}{g'(r)}\right] = \ln\left[\lim_{r\to 0}\frac{\frac{r}{\sqrt{r^2+(z-L)^2}}+0}{\frac{r}{\sqrt{r^2+(z+L)^2}}+0}\right] = \ln\left[\lim_{r\to 0}\frac{\sqrt{r^2+(z+L)^2}}{\sqrt{r^2+(z-L)^2}}\right] = \ln\left[\frac{z+L}{z-L}\right]$$

$$\Rightarrow \quad \lim_{r\to 0} V(r,z) = \frac{\lambda}{4\pi\epsilon_o}\left[\ln(z+L) - \ln(z-L)\right]$$

b. To find the component of the electric field along the z-axis (which we can tell from symmetry is the only component of the field on that axis), we take the negative derivative with respect to z:

$$E_z = -\frac{\partial V}{\partial z} = \frac{\lambda}{4\pi\epsilon_o}\left[\frac{1}{z-L} - \frac{1}{z+L}\right]$$

Plugging-in $\lambda = \frac{Q}{2L}$ for the uniform charge density gives us the electric field shown above.

[It should be noted that it is a better idea to perform the gradient operation before taking a limit (in this case $r \to 0$), because some information about the potential that is used in the electric field can be lost when the limit is taken first. In this particular case, symmetry ensured that the electric field would only point along the z-direction on the z-axis, so it was safe (and mathematically more expedient) to take the limit first.]

Electrostatic Potential Field of Long Straight Line of Uniform Density

As with the case of the electric field, the potential for a long ("infinite") line of charge is of importance to us. So the natural step to take is to simply use the result above, and take the limit as $L \to \infty$. But when we try to do this, we find that the limit diverges. This is puzzling, as the electric field comes out right from the gradient, and it doesn't diverge in the limit. The answer has to do with our choice of where the potential is chosen to be zero, and we can fix the problem by simply choosing another position for the zero potential. Let's choose the position for zero potential to be: $r = r_o, \;\; z = 0$. If we do this, we get a revised version of our potential function for the line segment:

$$V(r,z)=\frac{\lambda}{4\pi\epsilon_o}\left[\ln\left(\frac{\sqrt{r^2+(z-L)^2}+L-z}{\sqrt{r^2+(z+L)^2}-L-z}\right)-\ln\left(\frac{\sqrt{r_o^2+L^2}+L}{\sqrt{r_o^2+L^2}-L}\right)\right] \tag{2.3.3}$$

The reader can easily confirm that in fact $V(r_o,0)=0$. It's not immediately clear how this rescues our limit of $L\rightarrow\infty$. To see how this happens requires a bit of math, but it is useful to see, so here goes... First, it's clear that when the line becomes infinitely-long, the value of z is irrelevant, and might as well be set to zero. Plugging this in, and dividing the numerator and denominator of both natural log arguments by L simplifies our potential to:

$$V(r)=\frac{\lambda}{4\pi\epsilon_o}\left[\ln\left(\frac{\sqrt{1+\frac{r^2}{L^2}}+1}{\sqrt{1+\frac{r^2}{L^2}}-1}\right)-\ln\left(\frac{\sqrt{1+\frac{r_o^2}{L^2}}+1}{\sqrt{1+\frac{r_o^2}{L^2}}-1}\right)\right] \tag{2.3.4}$$

Now we make use of the following expansion to first order for very small ϵ:

$$\sqrt{1+\epsilon}\approx 1+\frac{1}{2}\epsilon \tag{2.3.5}$$

In our case, of course the ratios with L^2 in the denominator are very small, so we get:

$$V(r)=\frac{\lambda}{4\pi\epsilon_o}\left[\ln\left(\frac{1+\frac{r^2}{2L^2}+1}{1+\frac{r^2}{2L^2}-1}\right)-\ln\left(\frac{1+\frac{r_o^2}{2L^2}+1}{1+\frac{r_o^2}{2L^2}-1}\right)\right]=\frac{\lambda}{4\pi\epsilon_o}\left[\ln\left(4\frac{L^2}{r^2}+1\right)-\ln\left(4\frac{L^2}{r_o^2}+1\right)\right]=\frac{\lambda}{4\pi\epsilon_o}\ln\left(\frac{\frac{1}{r^2}+\frac{1}{4L^2}}{\frac{1}{r_o^2}+\frac{1}{4L^2}}\right) \tag{2.3.6}$$

Now take our limit of $L\rightarrow\infty$ at last, to obtain:

$$V(r)=\frac{\lambda}{4\pi\epsilon_o}\ln\left(\frac{r_o^2}{r^2}\right)=\frac{\lambda}{2\pi\epsilon_o}\ln\left(\frac{r_o}{r}\right) \tag{2.3.7}$$

We will find this result quite useful for cylindrical symmetries. It clearly still vanishes at $r=r_o$ and its negative gradient gives the correct field for the long line of uniform charge (Equation 1.3.21).

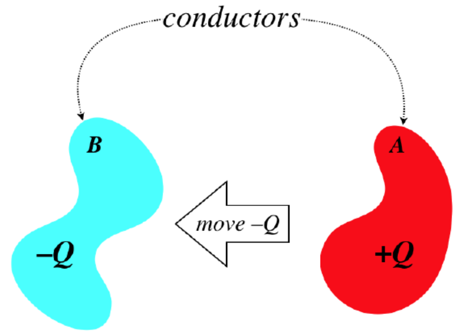
As with the case of electric fields, we can also compute potential scalar fields for non-uniform charge distributions. The reader is encouraged to go back to re-work Example 1.3.4 (for which the density was non-uniform) by starting with computing the electrostatic potential field.

2.4: Capacitance

Definition of Capacitance

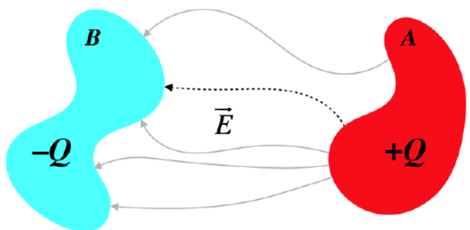
Imagine for a moment that we have two neutrally-charged but otherwise arbitrary conductors, separated in space. From one of these conductors we remove a handful of charge (say $-Q$), and place it on the other conductor.

Figure 2.4.1 – Charge Separated to Two Conductors



So now we have two conductors, one with a net charge of $+Q$, and the other with a net charge of $-Q$. Clearly there is an electric field pointing out of the former, and into the latter, with the field lines leaving and landing perpendicular to the surfaces.

Figure 2.4.2 – Field Between the Two Conductors

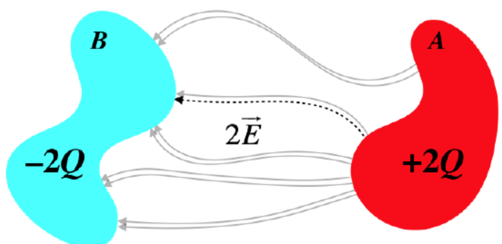


If we follow a field line leaving the positively-charged conductor and do a line integral along this field line until we reach the negatively-charged conductor, the result will be a decrease in electric potential, ΔV. That is, the positively-charged conductor will be an equipotential at a higher voltage than the equipotential that is the negatively-charged conductor.

$$V_A - V_B = \int_A^B \overrightarrow{E} \cdot \overrightarrow{dl} \tag{2.4.1}$$

Now let's imagine what happens if we take an additional $-Q$ from conductor A and move it over to conductor B. What would we expect to see change in the electric field? We would expect the *magnitude* of the electric field to change, but the field lines should be shaped exactly the same. This is because the relative densities of charge at different locations on each conductor shouldn't change, and the field lines still need to be perpendicular to the conducting surfaces.

Figure 2.4.3 – Doubling the Charge Separated Between the Conductors



Consequently, when we compute the potential change using the same path as before (i.e. following the same field line) with twice the charge, the only thing that should be different is the magnitude of the electric field at each point, and this magnitude should be exactly twice as great (e.g. we know this is true at each conducting surface where the field strength is proportional to the charge density). We express this fact that the potential difference across two conductors is proportional to the charge they separate in a simple equation:

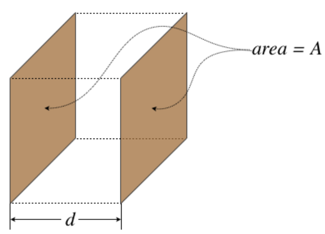
$$Q_{separated} \propto \Delta V \quad \Rightarrow \quad Q = C\Delta V \tag{2.4.2}$$

The constant C is called the *capacitance* of this two-conductor set-up. We associate this constant with the set-up because if the geometry is somehow changed (the conductors are pulled farther apart, one is rotated, the shape of one is altered, etc.), then the electric field lines from one to the other will change, and the line integral along one of those field lines will give a different potential difference, even though the same charge is separated. That is, the capacitance of a system of conductors is uniquely-defined by the physical structure of those conductors, but is unaffected by the amount of charge separated. The units of capacitance are obviously coulombs per volt, which is renamed for brevity to *farads*. A coulomb is a rather large amount of charge, and for most real-world applications of capacitance, we will see significantly less than a farad of capacitance, typically in the range of microfarads (μF).

Parallel-Plate Capacitor

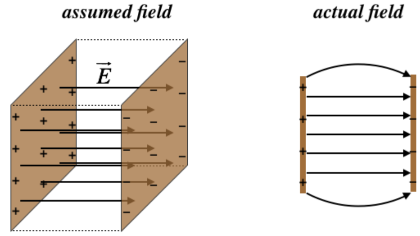
While capacitance is defined between any two arbitrary conductors, we generally see specifically-constructed devices called *capacitors*, the utility of which will become clear soon. We know that the amount of capacitance possessed by a capacitor is determined by the geometry of the construction, so let's see if we can determine the capacitance of a very simple capacitor – the *parallel-plate capacitor*.

Figure 2.4.4 – Parallel-Plate Capacitor



This kind of capacitor is modeled by two flat (obviously parallel) conducting plates, and while they are finite in extent, we approximate the fields between the plates with a uniform field. This approximation is quite good near the centers of the plates, but breaks down near the edges, where the field bows outward. This phenomenon (commonly referred to as *fringe effects*) plays a role that we will see later, but for now our approximation is that the field is uniform throughout the entire volume of the capacitor.

Figure 2.4.5 – Field Inside a Parallel-Plate Capacitor



While the capacitance depends only upon the structure of this capacitor, to figure out what the capacitance actually is, we need to place some charge on the plates, and compute the potential difference. We will then find that the ratio of these quantities is only a function of geometry. The field is uniform, which means that the field at the surface of one conductor is the same throughout the space between the conductors, and is the same at the other surface. But we already know what the field at the surface of a conductor is: $E = \frac{\sigma}{\epsilon_o}$. If there is a total charge separation of Q, then the uniform charge density is just that charge divided by the area of the conductor, giving:

$$E=\frac{\sigma}{\epsilon_o}=\frac{Q}{A\epsilon_o} \tag{2.4.3}$$

From the electric field, we can compute the potential difference. Taking a straight-line path from the positive plate to the negative plate, the path integral is easy, as the electric field is constant, and the angle between the field and displacement is zero throughout the path:

$$voltage\ drop=\Delta V=V_A-V_B=\int_A^B \vec{E}\cdot\vec{dl}=E\int_A^B dl=E\ d \tag{2.4.4}$$

Plugging this in above gives:

$$E=\frac{Q}{\epsilon_o A}=\frac{\Delta V}{d}\quad\Rightarrow\quad Q=\left(\frac{\epsilon_o A}{d}\right)\Delta V \tag{2.4.5}$$

The capacitance is the ratio of the charge separated to the voltage difference (i.e. the constant that multiplies ΔV to get Q), so we have:

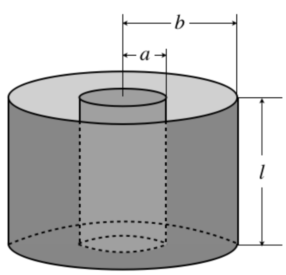
$$C_{parallel-plate}=\frac{\epsilon_o A}{d} \tag{2.4.6}$$

[*Note: From this point forward, in the context of voltage drops across capacitors and other devices, we will drop the "Δ" and simply use "V." That V represents a voltage difference and not an electrostatic potential at a point in space will be clear from context.*]

Coaxial Cylindrical Capacitor

Looking at the final answer for the capacitance of the parallel-plate capacitor, we see that indeed it only depends upon the structure of the conducting surfaces – in particular, the cross-sectional area and their separation. To see that this particular formula for capacitance is unique to parallel-plate capacitors, it is helpful to look at another conductor structure. One that is common to find in practice is two coaxial conducting cylinders.

Figure 2.4.6 – Cylindrical Capacitor



Once again we separated some charge across the conductors, and compute the potential drop between them. In this case, symmetry demands that the field is radial from the axis (it points outward if the inside conductor is the positively-charged one), although once again fringe effects occur near the edges, which we will ignore. We can use Gauss's law to show that the field is identical to that of a long line of charge with density λ:

$$E(r)=\frac{\lambda}{2\pi\epsilon_o r} \tag{2.4.7}$$

The line density is the charge per unit length, so in terms of the separated charge and the dimensions of the cylinder, have simply $\lambda\cdot l=Q$. As above, we can do a line integral from one plate to the other to get the voltage drop. Let's call the inner A and the outer cylinder B and assume the inner cylinder is positively-charged. Then we have:

$$V=\int_A^B \vec{E}\cdot\vec{dl}=\int_a^b E(r)\,dr=\frac{Q}{2\pi\epsilon_o l}\int_a^b \frac{1}{r}dr=\frac{Q}{2\pi\epsilon_o l}\ln\left(\frac{b}{a}\right) \tag{2.4.8}$$

Notice that as soon as we realized that the field within this capacitor is the same as that of a long line of charge, we could have simply used our previous result, Equation 2.3.7. Simply plug in $r=a$ for the potential of the smaller cylinder and $r=b$ for the potential of the larger cylinder, and find the difference of the two.

The capacitance of this set-up is the ratio of the charge to the potential drop, so we have:

$$C=\frac{Q}{V}=\frac{2\pi\epsilon_o l}{\ln\left(\frac{b}{a}\right)} \tag{2.4.9}$$

This is quite different from case of a capacitor with a parallel-plate geometry.

Work Functions of Metals

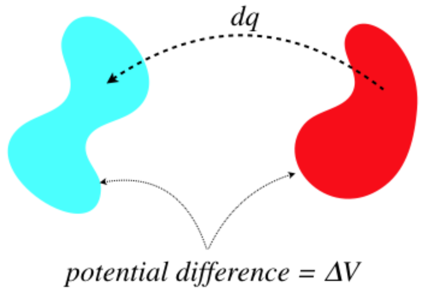
At this point it seems appropriate to ask: “Why don’t the oppositely-charged plates draw charges off one another? That is, why don’t sparks fly across the capacitor gap?” It turns out that whenever we get into the nitty-gritty details of how charges behave on an atomic scale, things get complicated very fast. You will examine a bit of this when/if you move on to Physics 9D, but for now we'll say that the charges are held on the conductor by the nearby charges in the metal (it’s the protons holding the electrons to the metal). The grip these nearby charges have is not infinite, and the strength of their grip is measured by a quantity known as the *work function* of the metal. This measures the amount of energy that must be added to a single electron to get it to escape the surface of the metal.

As we build charge on the two plates (and hold the geometry fixed), the electric field grows, increasing the force on the charges. If enough charge is added to the plates, then this force is sufficient to overcome the work function, and electrons spark across the gap. It should be pointed out that this maximum capacity for charge is *not* what is being referred-to when using the word “capacitance,” though of course they are related – all else being equal, a capacitor with a higher capacitance will hold more charge before sparking. This sparking condition can also work another way: If the potential across the two plates is held fixed, and the gap is shortened, then since $dV=-Edx$, the electric field gets stronger, and a spark can occur.

Energy Storage

What is the point of constructing capacitors? Energy storage. How do we know energy is stored in a capacitor? We take some charge away from one conductor and put it on the other, which means we are pulling charge away from opposite-sign charges, and pushing it toward same-sign charges. This requires putting in work, and accumulates electrical potential energy. We can calculate exactly how much energy is stored, and as always, we do so incrementally.

Figure 2.4.7 – Energy Accumulation in a Capacitor



When we move an infinitesimal charge dq across a potential ΔV, the increase in energy is the product of these values. But the potential difference can also be written in terms of the charge on the conductors and the capacitance, the latter of which is a constant so long as the geometry of the conductors is unchanged.

$$dU=dqV=dq\left(\frac{q}{C}\right) \quad \Rightarrow \quad U=\int_0^Q \frac{q\,dq}{C}=\frac{1}{2}\frac{Q^2}{C} \tag{2.4.10}$$

With the relation $Q=CV$, we can rewrite this expression of potential energy two other ways. To summarize:

$$U=\frac{1}{2}\frac{Q^2}{C}=\frac{1}{2}CV^2=\frac{1}{2}QV \tag{2.4.11}$$

The most commonly-used of these expressions in practice is the second one, since we are typically handed a capacitor of a known capacitance, and want to know how much energy we can store in it when we apply a known voltage. Measuring the charge is not generally an easy thing to do (except indirectly, knowing C and V).

Example 2.4.1

Imagine pulling apart two charged parallel plates of a capacitor until the separation is twice what it was initially. It should not be surprising that the energy stored in that capacitor will change due to this action. For the two cases given below, determine the change in potential energy. Also, provide a careful accounting of the energy: If the potential energy does down, explain where the energy goes, and if it goes up, explain where the energy comes from.

a. *the charged capacitor is not connected to anything that would allow it to change the charge on its plates*
b. *the charged capacitor is connected to a device that adjusts the charge on the plates, such that the plates of the capacitor are held at a constant electric potential difference*

Solution

For both cases, increasing the separation changes the physical structure of the capacitor, and since the capacitance only depends upon the physical structure (not the charge or voltage), we use the parallel-plate equation:

$$C=\frac{\epsilon_o A}{d}$$

Doubling the separation therefore reduces the capacitance to one-half its original value. We now use this fact for the two cases given.

a. Without the ability to adjust the amount of charge, it doesn't make sense to use $U=\frac{1}{2}CV^2$ to compute the potential energy change, because both variables are changing at once. We therefore use:

$$\left.\begin{array}{l} U_{before}=\dfrac{1}{2}\dfrac{Q^2}{C} \\ U_{after}=\dfrac{1}{2}\dfrac{Q^2}{\frac{1}{2}C} \end{array}\right\} \quad U_{after}=2U_{before}$$

So the electric potential energy within the capacitor doubles, but where does this energy come from? Well, the plates are oppositely-charged, so they attract each other. Pulling them apart requires exertion – ***work must be done on the system****. To compute the work done, we need to know the force required to barely get the plates to separate further (more force that this would accelerate the plates, which would bring unwanted kinetic energy into the calculation). This minimum force just equals the force of attraction between the plates, but what is this?*

Consider a single charge q on one of the plates. It is in a uniform field caused by the other plate, so it feels a force toward the other plate equal to:

$$F_{single\;charge}=qE_{other\;plate}$$

Of course, every charge on this same plate feels this force, so the total force on the plate is simply the sum of these forces:

$$F_{on\;plate\;A}=Q_{on\;plate\;A}E_{by\;plate\;B}$$

But the field due to the other plate is not the same as the total field within the capacitor – the total field is a superposition of the fields due to ***both*** *plates, and both plates contribute the same amount of field in the same direction, so in terms of the field in the capacitor, we have:*

$$F_{on\;plate\;A}=Q_{on\;plate\;A}\left[\frac{1}{2}E_{total}\right]=\frac{1}{2}QE$$

The charge doesn't change while the plates are pulled apart, so the electric field doesn't change either, which means that the force between the plates remains constant, making the work calculation easy. The amount of additional separation is d, so:

$$W=\int \overrightarrow{F}\cdot \overrightarrow{dl}=F\Delta x \quad \Rightarrow \quad W=\left(\frac{1}{2}QE\right)(d)$$

The quantity $E\cdot d$ is the original potential difference V. The energy put into the system by work is therefore $\frac{1}{2}QV$, which equals precisely the potential energy the system started with, confirming that the potential energy is doubled.

b. Holding the potential between the plates fixed suggests using a different equation to determine the effect on the potential energy:

$$\left.\begin{array}{l} U_{before}=\frac{1}{2}CV^2 \\ U_{after}=\frac{1}{2}\left(\frac{1}{2}C\right)V^2 \end{array}\right\} \quad U_{after}=\frac{1}{2}U_{before}$$

*This is a bit puzzling – clearly work must be done to separate the oppositely-charged plates, which adds energy to the system, but somehow the stored potential energy goes down?! The answer lies in the fact that to keep the potential the same as the plates separate, **charge must leave the plates**. Wherever this charge goes, it accumulation at that other place increases the potential energy there. So energy leaves the system and is stored as potential energy elsewhere.*

To determine how much energy is gained by the "other place," we multiply the charge moved (which we know is half the charge on the capacitor, since its capacitance goes down by that factor and the voltage stays the same) by the potential difference:

$$\Delta U_{other\ place}=Q_{moved}V=+\frac{1}{2}QV$$

This is the entirety of the energy that starts in the capacitor! But the energy doesn't end up with zero potential energy because there is work done in the separation. We can compute the work as we did in part (a), except that we need to keep in mind that the charge and electric field are changing as the plates are being separated. We know the relationship between the field and the charge, so with the voltage difference held constant, we have:

$$\left.\begin{array}{l} V=E\cdot x \\ E=\frac{\sigma}{\epsilon_o}=\frac{Q}{\epsilon_o A} \end{array}\right\} \quad F=\frac{1}{2}QE=\frac{1}{2}(\epsilon_o AE)E=\frac{\epsilon_o AV^2}{2x^2}$$

Now computing the work:

$$W=\int F\cdot dx=\int_d^{2d}\left(\frac{\epsilon_o AV^2}{2x^2}\right)dx=\frac{\epsilon_o AV^2}{2}\int_d^{2d}\frac{dx}{x^2}=+\frac{\epsilon_o AV^2}{4d}$$

Plugging in $C=\dfrac{\epsilon_o A}{d}$ shows that the work done is $\frac{1}{2}\left(\frac{1}{2}CV^2\right)$, which is half the original stored energy. So in showing that an amount equal to the starting energy exits the capacitor with the charge, and half the original energy enters the system via work, we have confirmed that the final potential energy is half of what it was at the beginning.

Energy Density in Electric Fields

An alternative way to discuss energy storage is in terms of the electric field. The simplest way to see this is to look at the energy stored in a parallel-plate capacitor:

$$U=\frac{1}{2}CV^2=\frac{1}{2}\left(\frac{\epsilon_o A}{d}\right)(Ed)^2=\frac{1}{2}\epsilon_o(Ad)E^2 \tag{2.4.12}$$

Notice that the quantity Ad is the volume of the parallel-plate capacitor. If we divide both sides of this equation by that volume, we get the *energy density* of the electric field, which we can express more generally (for any electric field, not just one within a parallel-plate capacitor):

$$u=\frac{1}{2}\epsilon_o E^2\ , \quad U=\int u\ dV \tag{2.4.13}$$

We interpret this as follows: At any point in space where an electric field is present, there is an electrical potential energy density given by the above expression. If we want to know how much *PE* is present in a finite volume of space, we need only integrate this density over the volume.

Note that we have only *claimed* that this works for all electric fields, while only deriving it for a parallel-plate capacitor. We can support this claim by demonstrating that it also works for the cylindrical capacitor. We know the electric field for this configuration is that of a line of charge, so we need to integrate the energy density derived from that field between the two cylinders of such a capacitor.

$$U=\int\left[\frac{1}{2}\epsilon_o E^2\right]dV \tag{2.4.14}$$

The infinitesimal volume element in cylindrical coordinates is given at the end of Section 1.3, and the limits of integration are straightforward, so plugging these in along with the function of the electric field (Equation 1.3.21):

$$\begin{aligned} U &= \frac{1}{2}\epsilon_o\int\left[\frac{\lambda}{2\pi\epsilon_o r}\right]^2[rdr\;d\phi\;dz] \\ &= \frac{\lambda^2}{8\pi^2\epsilon_o}\int_a^b\frac{dr}{r}\int_0^{2\pi}d\phi\int_0^l dz \\ &= \frac{\lambda^2}{8\pi^2\epsilon_o}\left[\ln\left(\frac{b}{a}\right)\right][2\pi]\,[l] \\ &= \frac{(\lambda\cdot l)^2}{4\pi\epsilon_o l}\ln\left(\frac{b}{a}\right) \\ &= \frac{1}{2}\frac{Q^2}{\frac{2\pi\epsilon_o l}{\ln\left(\frac{b}{a}\right)}} \end{aligned} \tag{2.4.15}$$

Comparing the denominator with Equation 2.4.9 shows that it is the capacitance, which then means that this quantity matches the energy stored according to Equation 2.4.11.

Example 2.4.2

Consider a solid conducting sphere of radius R which holds a total charge of Q on its surface. In Equation 2.1.8 we found that this system stores a potential energy of:

$$U=\frac{Q^2}{8\pi\epsilon_o R}$$

a. *Show that this is the potential energy stored in the electric field.*
b. *Find the capacitance of the sphere [we can treat the system as though there is another conducting sphere at $r=\infty$ to give us two conductors].*

Solution

a. The electric field is zero within the conducting material, so we need to integrate the energy density over the volume of all space from $r=R$ to $r=\infty$. The charge density is spherically symmetric, which means that the field looks identical to that of a point charge positioned at the center of the sphere (we can prove this easily using Gauss's law). The integral over the polar and azimuthal angles for this symmetric field give the usual factor of 4π, so all we need to do is the radial part of the integral:

$$U=\int\left[\frac{1}{2}\epsilon_o E^2\right]dV=\frac{1}{2}\epsilon_o\,(4\pi)\int_R^\infty\left[\frac{Q}{4\pi\epsilon_o r^2}\right]^2r^2dr=\frac{Q^2}{8\pi\epsilon_o}\int_R^\infty\frac{dr}{r^2}=\frac{Q^2}{8\pi\epsilon_o R}$$

b. Using our usual convention, the electrostatic potential at infinity is zero, so the potential difference between the two conductors (one of them a sphere at infinity) is simply the electrostatic potential at the surface of the sphere. The field there is identical to that of a point charge, so the potential difference is:

$$\Delta V = \dfrac{Q}{4\pi\epsilon_o R}$$

The capacitance is now simply:

$$C = \dfrac{Q}{V} = 4\pi\epsilon_o R$$

Note that the result depends only upon the geometry (not the charge or potential), as it should. We can double-check this result using what we found in part (a):

$$U = \dfrac{Q^2}{8\pi\epsilon_o R} = \dfrac{1}{2}\dfrac{Q^2}{C} \quad \Rightarrow \quad C = 4\pi\epsilon_o R$$

Example 2.4.3

In Section 2.1 we computed the energy stored in a sphere of uniformly-distributed charge of radius R, obtaining the result in Equation 2.1.13:

$$U = \dfrac{3Q^2}{20\pi\epsilon_o R}$$

Use the energy density of the electric field to confirm this result. [Note that the field within the sphere is not zero, and behaves differently than the field outside the sphere.]

Solution

We need to integrate the energy density over the volume to get the total potential energy. In this case, there are two regions to integrate over, because each has a different electric field. Let's address the field outside the sphere first. It is spherically symmetric, so the field outside it is identical to that of a point charge. Integrating from $r=R$ to $r=\infty$ gives the same result we obtained above for the conducting sphere:

\[U_{outside} = \dfrac{Q^2}{8\pi\epsilon_oR}\nonumber\]

Now for the region inside the sphere $r<R$ In this region, we can use Gauss's law to determine the field at a distance r from the center. Construct a gaussian surface with a radius r. The field is radial thanks to spherical symmetry, so it is perpendicular to the surface at all points on the gaussian surface, and it has the same magnitude everywhere on that surface.

$$\dfrac{q_{enclosed}}{\epsilon_o} = E\left(r\right)A = E\left(r\right)\left[4\pi r^2\right]$$

Now we need to find the enclosed charge. The density is uniform, so if we know that that is, we can simply multiply it (no need for an integral) by the volume of the gaussian surface. Well, we can determine this density using the total charge and total volume, so:

$$\rho = \dfrac{Q}{V} = \dfrac{Q}{\frac{4}{3}\pi R^3} \quad \Rightarrow \quad q_{enclosed} = \rho\left(\dfrac{4}{3}\pi r^3\right) = Q\left(\dfrac{r}{R}\right)^3$$

Plugging this back in above gives the electric field at all positions inside the sphere:

$$E\left(r\right) = \dfrac{q_{enclosed}}{4\pi\epsilon_o r^2} = \dfrac{Qr}{4\pi\epsilon_o R^3}$$

Now plug this into the energy density and integrate over the volume:

$$U_{inside} = \int\limits_{inside}\left[\dfrac{1}{2}\epsilon_o E^2\right]dV = 4\pi\int\limits_0^R \dfrac{1}{2}\epsilon_o\left[\dfrac{Qr}{4\pi\epsilon_o R^3}\right]^2 r^2 dr = \dfrac{Q^2}{8\pi\epsilon_o R^6}\int\limits_0^R r^4 dr = \dfrac{Q^2}{40\pi\epsilon_o R}$$

If we add the potential energy stored inside the sphere to the potential energy stored outside, we get the desired answer:

$$U_{total} = \frac{Q^2}{8\pi\epsilon_o R} + \frac{Q^2}{40\pi\epsilon_o R} = \frac{3Q^2}{20\pi\epsilon_o R}$$

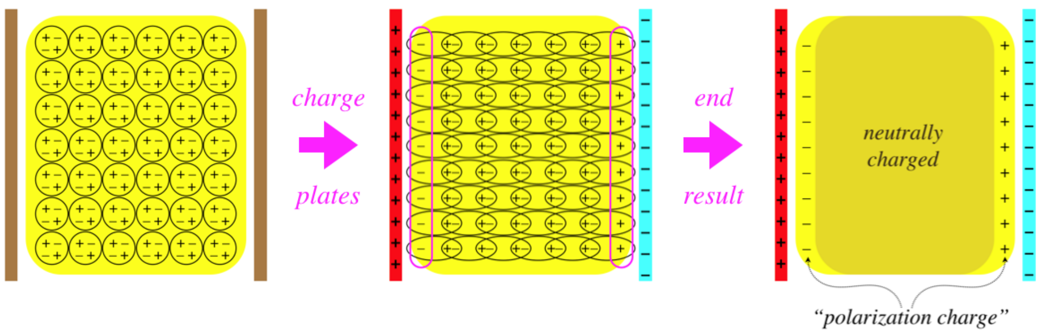
2.5: Dielectrics

Polarization

Up to now, we have placed all substances into one of two categories – insulators and conductors – distinguished by whether they hold charges utterly fixed or allow them completely free movement. Well, it probably isn't surprising that in reality substances generally fall between these two extremes. For now we will focus on the insulator side of the spectrum. Imagine a substance that doesn't allow charges to go wherever they please, but does allow for the atoms or molecules binding these charges to morph their shapes. Such substances are called *dielectrics*, and they actually provide an effect similar to what is seen in conductors, though it is not extreme enough to completely cancel the field.

Start with a slab of neutrally-charged dielectric located between two neutrally-charged conductor plates. If the plates are then charged, the electric field produced between the two plates pull the charges in the dielectric in opposite directions. Within the dielectric, the positive and negative charges just pair-off differently, leaving a continued neutral charge. But on the surfaces, the separated charges don't pair-off with opposites, leaving a net charge on the two surfaces of the dielectric called *polarization charge*.

Figure 2.5.1 – Creation of Polarization Charge on a Dielectric



We saw this same thing happen in a conductor, but because the charges were totally free to move, they continued polarizing until the *net* field vanished within the conductor. In the case of dielectrics, the charges stop shifting long before the field of the polarization charge can cancel the field of the free charge, which means that there is still a net field remaining within the dielectric at the end.

For the parallel-plate geometry in the figure above, the net field is easy to compute from the free and polarization charges, as they are both planes. We can similarly solve for the net field in the case of a dielectric inside a capacitor of concentric conducting cylinders. But things get far too complicated when the surface of the dielectric is not orthogonal to the external field, so we will only consider these simpler geometries. Furthermore, we will assume that the entire dielectric is the same material – the amount that the charges are able to separate depends upon the molecules, so they have to be the same throughout the sample.

With these restrictions in place, we can conclude that the field caused by the polarization charge (called the *polarization field*) is in the direction opposite to the applied field, and since the applied field is always stronger, we can write:

$$\left|\overrightarrow{E}_{total}\right| = \left|\overrightarrow{E}_{applied}\right| - \left|\overrightarrow{E}_{polarization}\right| \tag{2.5.1}$$

It's clear that increasing the strength of the applied field pulls harder on the charges in the dielectric, and should increase the polarization charge. We make the further assumption (demonstrated experimentally, as long as the applied field is not too strong) that if we double the field strength, the polarization field also doubles. That is, the polarization field is proportional to the applied field. Combining this with the equation above means that the applied field is proportional to and in the same direction as the total field (with the applied field stronger), and we will write the constant of proportionality, called the *dielectric constant* as a lower-case Greek letter kappa:

$$\kappa \equiv \frac{\left|\overrightarrow{E}_{applied}\right|}{\left|\overrightarrow{E}_{total}\right|} \tag{2.5.2}$$

Note that this constant is dimensionless, is greater than or equal to 1. It is equal to 1 for a vacuum (where there are no charges to polarize), or a perfect insulator (which allows no charge movement at all).

Effects on Capacitors

The most common application of dielectrics is in capacitors, as one would guess from the figure. How is the capacitance affected by the presence of this substance? Given the same charges on the plates, the polarization charge reduces the electric field between the plates compared to the vacuum case, so the voltage difference is decreased. With a smaller voltage for the same charge on the plates, the capacitance is *increased*. Specifically, it is increased by a factor of exactly the dielectric constant:

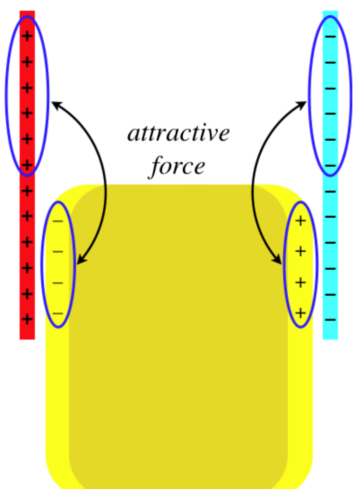
$$C_{dielectric} = \frac{Q_{on\ plates}}{\Delta V_{total}} = \frac{Q_{on\ plates}}{\int\limits_{plate\ A}^{plate\ B} \overrightarrow{E}_{total} \cdot \overrightarrow{dl}} = \frac{Q_{on\ plates}}{\int\limits_{plate\ A}^{plate\ B} \frac{1}{\kappa}\overrightarrow{E}_{applied} \cdot \overrightarrow{dl}} = \kappa \frac{Q_{on\ plates}}{\Delta V_{vacuum}} = \kappa C_{vacuum} \tag{2.5.3}$$

We noted several sections ago that the primary purpose of a capacitor is to store electrical potential energy. Let's now consider what happens to the potential energy when a dielectric is added into or taken out of a capacitor. Adding a dielectric increases the capacitance, and taking it away reduces it. From here, we can follow the calculations performed in Example 2.4.1. It was noted there that the change in energy depends upon what is held constant as the capacitance is changed – the charge on the plates, or the potential difference, and that must be taken into account here as well. The only difference here is that the capacitance changes as a result of the dielectric constant changing, rather than a change in the separation of the plates.

The overall result is the same – with the capacitance increasing when the dielectric is inserted, the potential energy goes up if the potential difference is held fixed, and it goes down if the plates are forces to keep the same charge. But in the example cited, the energy changes were accounted-for by considering the work done in separating the plates. Here the plate separation doesn't change, so if there is no work done, how can we account for where the energy comes from or goes to?

Well, in fact there *is* work done in the removal or insertion of the dielectric. We can see this by looking at what the system must look like when the dielectric is partially-inserted. The polarization charge on the surface of the dielectric that is between the plates will be attracted to the free charge on the part of the plates that are still separated by vacuum:

Figure 2.5.2 – Force on a Partially-Inserted Dielectric



In order to pull the dielectric out of the capacitor requires that work be added to the system (equivalent to increasing the plate separation in Example 2.4.1), while allowing the dielectric to be pulled into the capacitor removes energy from the system in the form of work done on the dielectric. This analysis can be performed "in reverse" to determine the force exerted on a partially-inserted dielectric by the capacitor. In Physics 9A, we learned that the force due to a potential energy field is equal to the negative of the gradient of the potential energy (see Physics 9A Libretext, Section 3.6):

$$\overrightarrow{F} = -\overrightarrow{\nabla} U \tag{2.5.4}$$

The change only occurs parallel to the plates, which we will call the y-direction, so this simplifies to just one component:

$$F_y = -\frac{dU}{dy} \tag{2.5.5}$$

When the dielectric moves into the plates an additional tiny distance dy, the potential energy of the system changes. How much it changes depends once again upon whether the charge on the plates or the potential difference remains constant during the process (the dependence of work done on which quantity is held constant was also a feature of Example 2.4.1). So all one needs to do is write down the potential energy for the capacitor at whatever position the dielectric is in, recalculate it for the dielectric inserted an additional distance dy, take the difference to obtain dU, then divide by dy. An important part of this process is noting that the capacitor with a partial dielectric inserted is equivalent to two separate capacitors, one with a vacuum between the plates, and one with dielectric between them. The total energy of the system is the sum of the energy in these two capacitors, and one needs to keep in mind that as each plate is an equipotential, the potential difference between the two plates for the two separate capacitors is the same.

Permittivity

One way that the dielectric constant is accounted-for is within another constant that we are already familiar with. To see this, consider how the capacitance of a parallel plate capacitor containing a vacuum changes when a dielectric is inserted:

$$C_{vacuum} = \frac{A\epsilon_o}{d} \quad \Rightarrow \quad C_{dielectric} = \kappa\frac{A\epsilon_o}{d} = \frac{A\epsilon}{d}\ , \quad where: \ \ \epsilon \equiv \kappa\epsilon_o \tag{2.5.6}$$

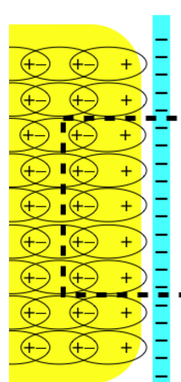
At last it's clear why the 'o' subscript was used up to now: The 'o' refers to the vacuum, which is why it is called the permittivity of *free space*. The quantity ϵ (with no subscript) is simply called the *permittivity* of the dielectric. The advantage of making this definition is that it saves us the trouble of re-deriving all the results where we used ϵ_o previously for cases where a dielectric medium is involved. It turns out that we can simply blindly replace the free-space constant with that for the dielectric, and all the same results apply. There is, however, one important detail to keep in mind here, however.

We introduced the dielectric constant and then the permittivity as a means of ignoring the polarization charge. That is, the capacitance with a dielectric still satisfies $Q = CV$, where Q is the charge *on the plates, not the combination of the charge on the plates with the polarization charge*. Wherever we use the permittivity, the requirement that we only account for the *free charge* (the charge present that excludes the polarization charge) must be observed. An important example of following this requirement follows.

Gauss's Law in Media

Consider the case of employing Gauss's law to determine the electric field near the surface of a conducting plane, as we did in Figure 1.7.2, but this time with a dielectric medium present outside the conducting surface.

Figure 2.5.3 – Gaussian Surface for a Conducting Surface Near a Dielectric



This gaussian surface encloses both the free charge on the conducting surface and the polarization charge on the surface of the dielectric. The net flux out of the gaussian surface (all of it gong through the left side of the surface shown) is lower than it would be without the dielectric, because the polarization charge cancels some of the free charge. The difference in flux comes entirely from the difference in the electric field, which we already know how to express:

$$\Phi_{with\ dielectric} = E_{total}A = \frac{1}{\kappa}E_{applied}A = \frac{1}{\kappa}\Phi_{without\ dielectric} = \frac{\epsilon_o}{\epsilon}\Phi_{without\ dielectric} \tag{2.5.7}$$

According to Gauss's law, the flux without the dielectric is just $\frac{Q_{plate}}{\epsilon_o}$, so we can express Gauss's law in terms of the free charge enclosed rather than the total charge enclosed using the dielectric permittivity:

$$\oint \overrightarrow{E}\cdot d\overrightarrow{A} = \frac{Q_{free}}{\epsilon} \tag{2.5.8}$$

In local form, this becomes:

$$\overrightarrow{\nabla}\cdot\overrightarrow{E} = \frac{\rho_{free}}{\epsilon} \tag{2.5.9}$$

All of the other appearances of the permittivity that we have seen similarly carry-over, most notably:

$$u = \frac{1}{2}\epsilon E^2 \tag{2.5.10}$$

Example 2.5.1

A point charge is held fixed in a medium with a dielectric constant equal to 2 near a large conducting plane. If the dielectric is now removed, describe how the following quantities change:

a. *the force on the point charge by the conductor*
b. *the charge induced on the surface of the conductor*

Solution

a. The electric field is the same in both cases, with the exception of the value of the permittivity, which is twice as great when the dielectric is in place than when it is not. This weakens the electric field of the point charge by a factor of 2. The induced charge on the conducting surface therefore responds by producing an equivalent field as if originating from an image charge. This weaker induced field results in a force on the point charge that is half as strong as when the dielectric is absent.

b. The charge induced on the surface of the conductor equals the negative of the point charge whether the dielectric is present or not. We can prove this a couple of ways. The simplest is to note that Newton's third law requires that if the force on the point charge is half as much with the dielectric, then the force on the conductor is also half as great. But the field of the point charge is half as strong, so the charge on which this field is acting must not be changed.

A second way to show this is to note that the electric field at the surface of a conductor in terms of the charge density is:

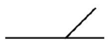
$$E = \frac{\sigma_{free}}{\epsilon}$$

We already know that the field is half as strong with the dielectric in place, and since $\epsilon = 2\epsilon_o$, the charge density must be the same in both cases.

2.6: Static Networks

Symbolic Diagrams

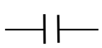
We have discussed capacitors as theoretical constructs, but in fact they are common electrical components, used in many devices. One typical use is as a source of a “burst” of electrical energy that is stored when a voltage source charges it at its leisure, but then is unleashed more quickly to do some job that the original voltage source cannot supply so fast. We will see more uses as well, in a later chapter. To discuss capacitors in this practical context, it becomes helpful to introduce some symbolic notation. We start with four symbols:

equipotential: 

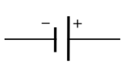
Solid, unbroken lines that connect components represent equipotentials. These can consist of single lines, or lines that branch-off from each other. While it is usually harmless to do so, thinking of these lines as conducting wires in the physical world can sometimes be dangerous when they are first encountered, as we will use this same set of symbols later when we discuss electric current, where wires are not equipotentials. Also, problems are easier to solve if you keep in mind that everything connected by straight lines are at the same potential, rather than just a conduit of charge flow.

switch: 

This is a component that facilitates talking about processes that involve connecting and disconnecting other components. When the switch is closed, it simply becomes an equipotential.

capacitor: 

The symbol looks like the side view of a parallel-plate capacitor, but it can represent any geometry of capacitor, with or without a dielectric within it. Note that the connection of a "plate" with a straight line suggests that they are at the same potential (and they are), but because there is a gap between the two plates, they are of course not at the same potential. This is the first of many components we will encounter that generally involve a potential difference from one side to the other.

battery: 

Up to now, we have talked about "holding a capacitor at a constant potential difference," while we do such things as pull the plates apart or insert a dielectric. We will no longer require this cumbersome language, as this is precisely the function of a battery. When charges move into or out of a capacitor, the potential difference (or *voltage*) across the plates changes, but the two "plates" of the battery shown in the symbol remain at the same potential difference at all times. If one side of the battery is connected by an equipotential to one side of a component, then that side of the component remains at the potential provided by the battery. If maintaining this potential requires charge movement, the battery supplies or accepts the charge as needed. [*It should be noted that there are several symbols one may encounter for batteries. The '–" and '+' in the symbol shown are actually superfluous and don't always appear – the larger "plate" is always the one at higher potential. Also, a symbol with more than just two "plates" (alternating in size) is often used for a battery.*]

Equivalent Capacitance

Next we will combine multiple components together, connected by equipotentials. A *circuit* is a closed-loop of components and equipotentials. Often these systems of components involve branching equipotentials, which results in many closed loops. Such a system is a combination of many circuits, and often the entire system referred to collectively as a circuit as well, but a more precise term for such a system is *network*. The reader is likely to encounter both terms to describe such systems in this work.

In order to analyze these systems, it is useful to develop some tools for examining circuit fragments of multiple capacitors (each of which can have a different capacitance). The simplest such fragment is one in which they are connected "consecutively," as shown in the figure below.

Figure 2.6.1 – Capacitors in Series

The color-coded diagram on the left emphasizes the equipotentials and plates that are all at the same potential – the left plate of capacitor #1 is at potential V_A, the right plate of capacitor #1 *and* the left plate of capacitor #2 are at potential V_B and the right plate of capacitor #2 is at potential V_C. The voltage across the capacitors are therefore:

$$V_1 = V_A - V_B \;, \qquad V_2 = V_B - V_C \tag{2.6.1}$$

This means that the voltage difference between the equipotentials on the two ends of the combination of capacitors is simply the sum of the voltages across the capacitors:

$$V_{tot} = V_A - V_C = (V_A - V_B) + (V_B - V_C) = V_1 + V_2 \tag{2.6.2}$$

What about the charge on each capacitor? Well, we know that the plates on a single capacitor have equal charge with opposite signs. The blue segment is isolated, so cannot receive any outside charge, which means that whatever charge is on the right plate of capacitor #1, the same amount of charge with the opposite sign is on the left plate of capacitor #2. Therefore the charge on each capacitor in this configuration is the same.

Whenever two or more capacitors are arranged in this way, such that they satisfy the above properties (the voltage across the combination is the sum of the voltages of each one, and the charges are the same on all of them), we say that these capacitors are *in series*. We have a total voltage difference for the combination, and a single amount of charge on a plate, so we can define an *equivalent capacitance* for the arrangement. Specifically:

$$\left.\begin{array}{l} Q = C_{eq}V_{tot} \\ Q = C_1V_1 \\ Q = C_2V_2 \\ V_{tot} = V_1 + V_2 \end{array}\right\} \quad \frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} \quad \text{(series)} \tag{2.6.3}$$

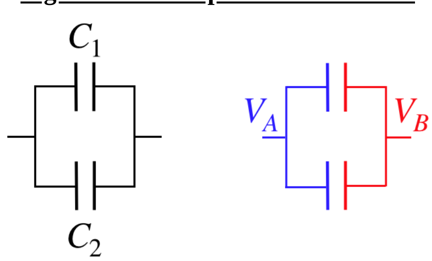
Note that if there are more than two capacitors in series, we only need to add additional inverse terms.

Alert

It is not uncommon to mistakenly declare capacitors to be in series, simply because they "look consecutive," when in fact they are not in series. It is important to make sure that the criteria involving voltage and charge are in effect before computing an equivalent series capacitance.

Another basic arrangement of capacitors reverses the roles of charge and voltage, and is shown in the figure below.

Figure 2.6.1 – Capacitors in Parallel



Once again the right diagram color-codes the equipotentials, and this time we can see that the two capacitors have the same potential difference across them:

$$V_1 = V_A - V_B = V_2 \tag{2.6.4}$$

As for the charge, each plate must get charge from outside (through the equipotentials), and different amounts of charge will go to the capacitors, and the total charge supplied to this system will equal the sum of the charges supplied to each individual capacitor. When these two criteria hold, we say that the capacitors are *in parallel*.

Alert

This is another warning about declaring two capacitors to be in parallel, simply because they look like it. The simplest test for parallel capacitors is to check to see if an equipotential directly connects both plates of one capacitor to the corresponding plates of the other capacitor.

For an equivalent capacitance, we put together the two new criteria, and get a new relation:

$$\left.\begin{aligned} Q_{tot} &= C_{eq}V \\ Q_1 &= C_1V \\ Q_2 &= C_2V \\ Q_{tot} &= Q_1+Q_2 \end{aligned}\right\} \quad C_{eq}=C_1+C_2 \quad \text{(parallel)} \tag{2.6.5}$$

If there are more than two capacitors in parallel, then of course the equivalent capacitance is the sum of all the individual capacitances.

Example 2.6.1

Show that for both the series and parallel cases, the energy stored in the equivalent capacitor equals the sum of the energies in the individual capacitors.

Solution

For the series case, the charge is the same on both capacitors, so the total stored energy is:

$$U=U_1+U_2=\frac{Q^2}{2C_1}+\frac{Q^2}{2C_2}=\frac{Q^2}{2}\left(\frac{1}{C_1}+\frac{1}{C_2}\right)=\frac{Q^2}{2C_{eq}}$$

For the parallel case, the voltage is the same across both capacitors:

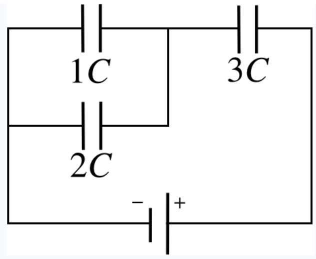
$$U=U_1+U_2=\frac{1}{2}C_1V^2+\frac{1}{2}C_2V^2=\frac{1}{2}(C_1+C_2)V^2=\frac{1}{2}C_{eq}V^2$$

Networks of Capacitors

Now just because we only have an equation for a pair of capacitors, it doesn't mean that we can only solve problems with two capacitors. We can solve much bigger networks, in a bootstrap manner by combining pairs of capacitors to form equivalent capacitors, then treating equivalent capacitors as if they are "real" capacitors, and combining *them* into new equivalent capacitors. Once enough reductions have occurred, one can conclude how much charge comes off the battery, and then "work backwards," keeping in mind that the amount of charge on an equivalent capacitor is the same as the charge on its constituent capacitors if they are in series, and the voltage difference across the equivalent capacitor is the same as across its constituent capacitors if they are in parallel. Once all the equivalent capacitors have been unwound, the charges and voltages (and therefore the energies as well) are known for every capacitor in the network. Mastering this process is only a matter of practice, so here's an example...

Example 2.6.2

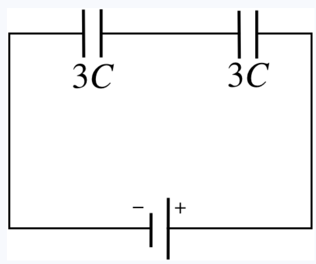
For the network in the figure, compute the fraction of the total energy supplied by the battery that goes to each of the individual capacitors.



Solution

We will begin by calling the voltage across the battery 'V'. It will divide out of the final answer, but we need something to put into the algebra. Now we begin the process of combining the capacitors to create a single equivalent capacitance for the network.

It might seem like $1C$ and $3C$ are in series, but they are not! The equipotential that connects to the left plate of $3C$ splits and connects to the right plates of two capacitors. This means that the sum of the charges on the right plates of $1C$ and $2C$ is the negative of the charge on the left plate of $3C$. With the charges not equal for $1C$ and $3C$, they cannot be in series. On the other hand, tracing the equipotential attached to the left plate of $1C$ goes to the left plate of $2C$, while tracing the equipotential attached to the right plate of $1C$ goes to the right plate of $2C$, so they have equal voltage drops and are in parallel. We replace them with an equivalent capacitance (which is just the sum of their capacitances), and redraw the diagram:



These capacitors are clearly now in series, so we combine them to make a single equivalent capacitor:

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} = \frac{2}{3C} \quad \Rightarrow \quad C_{eq} = \frac{3}{2}C$$

With a single equivalent capacitor attached to the battery, we can compute how much charge leaves the battery, and the amount of energy supplied by the battery:

$$Q_{tot} = C_{eq}V = \tfrac{3}{2}CV$$
$$U_{tot} = \tfrac{1}{2}C_{eq}V^2 = \tfrac{3}{4}CV^2$$

Now comes the tricky part – the "unwinding." We start with the charge supplied by the battery. Whatever positive charge leaves the battery collects on the right plate of $3C$ - it can't go anywhere else. This tells us immediately the voltage across $3C$, and from that, the energy stored on it.

$$Q_3 = Q_{tot} = \tfrac{3}{2}CV \quad \Rightarrow \quad V_3 = \frac{Q_3}{3C} = \tfrac{1}{2}V$$
$$U_3 = \tfrac{1}{2}(3C)V_3^2 \quad \Rightarrow \quad U_3 = \tfrac{3}{8}CV^2$$

The equivalent capacitor for $1C$ and $2C$ is in series with $3C$, so the sum of their voltages must equal the total voltage across the combination, which is just the voltage of the battery. Therefore, unsurprisingly, the voltage across equivalent capacitor for $1C$ and $2C$ is also $\frac{1}{2}V$. The two individual capacitors in this parallel combination have the same voltage differences as the voltage difference of their combination, so we can compute the energy stored in each of these (we don't need to compute the charge on each capacitor, but we could easily do so, if we wanted)

$$U_1 = \frac{1}{2}(1C)CV_1^2 = \frac{1}{2}(1C)C\left(\frac{1}{2}V\right)^2 = \frac{1}{8}CV^2$$
$$U_2 = \frac{1}{2}(2C)CV_2^2 = \frac{1}{2}(2C)C\left(\frac{1}{2}V\right)^2 = \frac{1}{4}CV^2$$

The fractions of energy stored in each capacitor are therefore:

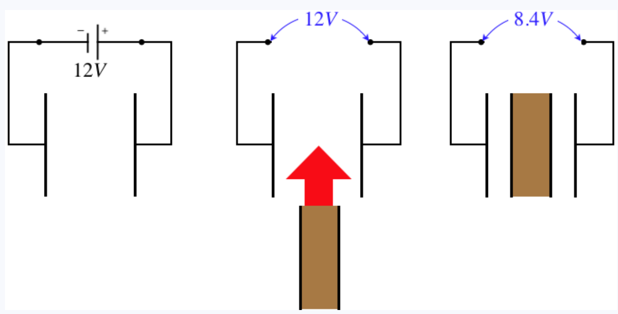
$$\frac{U_1}{U_{tot}} = \frac{\frac{1}{8}CV^2}{\frac{3}{4}CV^2} = \frac{1}{6}$$
$$\frac{U_2}{U_{tot}} = \frac{\frac{1}{4}CV^2}{\frac{3}{4}CV^2} = \frac{1}{3}$$
$$\frac{U_3}{U_{tot}} = \frac{\frac{3}{8}CV^2}{\frac{3}{4}CV^2} = \frac{1}{2}$$

Note that these fractions add to 1, which confirms that the total energy that leaves the battery equals the sum of the energies in the three capacitors.

There are other applications of these tools that don't neatly fall into this template. These sorts of examples require some thought as to whether series or parallel applies. Here are a couple of the slightly-more-offbeat examples...

Example 2.6.3

A parallel-plate capacitor with a vacuum between its plates is charged by connecting it to a $12V$ battery, and after it is fully charged, the battery is disconnected. Next an insulator with a thickness equal to half the width of the gap in the original capacitor is sandwiched between two thin conducting plates and is inserted between the plates of the original capacitor. After this is done, the voltage across the full system is measured to be $8.4V$. Find the dielectric constant of the insulator.



Solution

Placing the dielectric between the plates creates three separate regions within the capacitor. Thanks to the electric field being perpendicular to the plates, each end of the dielectric is an equipotential, which means that we can treat the three regions as though they are separate parallel-plate capacitors which are in series. The center capacitor has a dielectric while the outer capacitors have vacuum gaps. For the sake of doing the math, we'll call the gap size of the original capacitor $2d$ (making the thickness of the dielectric d and the thicknesses of the outer two capacitors $\frac{1}{2}d$. We'll call the area of the plates A, and the dielectric constant (which we seek) κ.

a. The capacitance of the original capacitor is:

$$C_{before} = \frac{\epsilon_o A}{2d}$$

Treating the new configuration as three capacitors in series and computing the equivalent capacitance gives:

$$\frac{1}{C_{after}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} = \frac{\frac{1}{2}d}{\epsilon_o A} + \frac{d}{\kappa\epsilon_o A} + \frac{\frac{1}{2}d}{\epsilon_o A} \quad \Rightarrow \quad C_{after} = \frac{\epsilon_o A}{d}\left[\frac{\kappa}{\kappa+1}\right]$$

With the plates disconnected from the battery, the charge on the capacitor plates remains constant through this process (it has nowhere to go!). Setting the charges equal gives us a ratio of the before and after capacitances in terms of the ratio of the given voltages:

$$Q = C_{before}V_{before} = C_{before}V_{before} \quad \Rightarrow \quad \frac{C_1}{C_2} = \frac{V_2}{V_1} \quad \Rightarrow \quad \frac{\frac{\epsilon_o A}{2d}}{\frac{\epsilon_o A}{d}\left[\frac{\kappa}{\kappa+1}\right]} = \frac{\kappa+1}{2\kappa} = \frac{8.4V}{12V} \quad \Rightarrow \quad \kappa = 2.5$$

Example 2.6.4

Two capacitors are separately connected to batteries until they are fully charged, and then are disconnected. At this point the capacitors hold equal charge, but capacitor #1 stores three times as much energy as capacitor #2. The two capacitors are then connected to each other such that the positive lead of one capacitor is connected to the positive lead of the other, and likewise with the negative leads.

a. *In which direction does charge flow when this connection is made, and what fraction of the charge flows out of the capacitor that loses charge?*
b. *If the voltage difference between the positive plates and negative plates after the capacitors are connected is V, find the voltage across each capacitor before they were connected in terms of V.*
c. *Find the fraction of energy change in the system when the charges rearrange themselves after the capacitors are connected. Is it a gain or a loss?*

Solution

a. Start with the fact that both capacitors had equal charge, that we will call Q. Then apply the fact that capacitor #1 stored three times as much energy:

$$U_1 = 3U_2 \quad \Rightarrow \quad \frac{Q^2}{2C_1} = 3\frac{Q^2}{2C_2} \quad \Rightarrow \quad C_2 = 3C_1$$

When they are connected together, it is in parallel, because the plates of one capacitor are connected to their counterparts on the other capacitor with equipotentials. This means their plates are now forced to be at the same potential difference. Two capacitors with equal voltage differences will hold charges proportional to their capacitances:

$$\left.\begin{array}{l} Q_1 = C_1 V \\ Q_2 = C_2 V \end{array}\right\} \quad \frac{Q_1}{Q_2} = \frac{C_1}{C_2}$$

So capacitor #2 will have 3 times as much charge on its plates as capacitor #1 when the charge stops rearranging itself. They started with equal charge, so for capacitor #2 to have 3 times as much charge as #1, #1 must have lost half its starting charge. In terms of our defined value Q, capacitor #1 ends up with $\frac{1}{2}Q$, while capacitor #2 ends up with $\frac{3}{2}Q$.

b. We have the before and after conditions, and comparing them gives our answers:

$$\left.\begin{array}{ll} before: & Q = C_1V_1 \\ after: & \frac{1}{2}Q = C_1V \end{array}\right\} \quad V_1 = 2V$$

$$\left.\begin{array}{ll} before: & Q = C_2V_2 \\ after: & \frac{3}{2}Q = C_2V \end{array}\right\} \quad V_2 = \frac{2}{3}V$$

c. Writing the total energies before and after in terms of Q and V gives:

$$\begin{array}{ll} before: & U = \frac{1}{2}QV_1 + \frac{1}{2}QV_2 = \frac{1}{2}Q\left(2V\right) + \frac{1}{2}Q\left(\frac{2}{3}V\right) = \frac{4}{3}QV \\ after: & U = \frac{1}{2}Q_1V + \frac{1}{2}Q_2V = \frac{1}{2}\left(\frac{1}{2}Q\right)V + \frac{1}{2}\left(\frac{3}{2}Q\right)V = QV \end{array}$$

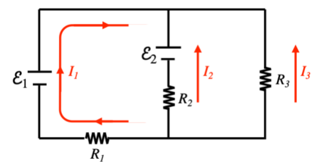
The system drops to three quarters of its original energy, so it loses one-fourth of what it started with.

Where does the energy go? Just for the sake of argument, let's suppose the capacitors are of the parallel-plate variety, and that capacitor #2 has three times the capacitance of capacitor #1 because its plates have three-times the area. With equal charges on each capacitor, this means that the charge density on capacitor #1 is three times greater than on capacitor #2. The fact that the charges are squeezed together more tightly on capacitor #1 is the reason that more potential energy is stored there. Allowing the charges to redistribute uniformly by connecting the capacitors results in the system evolving to a lower potential energy state. In terms of forces, the repulsion of charges on capacitor #1 exceeds the repulsion of those charges from capacitor #2, so net work is done by the electrical force in reaching the new state (reducing the potential energy), and since we assume that the kinetic energy given to the charges is ultimately dissipated away, the energy in the system at the end is lower than at the start.

CHAPTER OVERVIEW

3: DIRECT CURRENT CIRCUITS

We finally discard our requirement that charge remains static, and examine the effects associated with moving charge.



3.1: MOVING CHARGE

We begin with some mathematical tools for dealing with moving charge, keeping in mind the observed physical law that charge is neither created nor destroyed.

3.2: RESISTANCE AND ENERGY DISSIPATION

We now discard our assumption from electrostatics that conductors allow totally free (instantaneous) movement of electric charge, and take into account the effects of "electrical friction."

3.3: NETWORKS OF BATTERIES AND RESISTORS

We can employ analysis similar to what we used in static circuits to circuits that contain electric current passing through resistors, though it requires a twist to the notion of potential difference used in the static case.

3.4: KIRCHHOFF'S RULES

Not every conceivable circuit can be analyzed using the tools of the previous section. Here we learn some tools that can be used in more general cases.

3.5: RC CIRCUITS

Up to now, we have only considered the role of capacitors under static circumstances. We now incorporate them into our moving-charge networks.

3.1: Moving Charge

Electric Current

Up to now we have avoided talking about the details of moving charge, but we avoid it no longer. We begin by defining a quantity we will be using a lot – *electric current*. Simply put, this is the amount of charge that passes a fixed point in a given period of time:

$$I \equiv \frac{dq}{dt} \tag{3.1.1}$$

This has units of coulombs per second, which is given its own name: *amperes* or *amps*.

First off, we need to say that it is the *electrons* that do the moving – the protons are fixed in the nucleus of the atoms that are fixed in a lattice that constitutes the conductor. This may cause some confusion at first, since electrons are defined to have negative charge, and the current is defined to be in the direction of positive charge flow. This means that while electrons are moving in one direction, the current associated with this charge flow is in the opposite direction.

Digression: Charge Carriers

*There are other types of electrical current besides electrons moving through conductors. These currents are effectuated by other types charge carriers. One common variety of charge carrier is an ion, which is an atom that is not neutrally charged because it is either missing an electron or has an extra electron. This sort of charge carrier is most prevalent in biological systems, in fluids called electrolytes. Another charge carrier is not a true particle at all, but rather the **absence** of an electron (so it is positively-charged), called a hole. These are important in semiconductor physics, and come into play in the common electrical components of diodes and transistors.*

Second, we will discard the notion that these electrons are *completely* free to move within a conductor, as they actually will encounter something very similar to air resistance. If you recall from Physics 9A, air resistance is a dissipative force that comes about because particles that comprise the air collide with, and thereby transfer momentum to, the object experiencing moving through the air. A simple model of resistance in a conductor has the electrons colliding with the fixed atom nuclei. This is of course oversimplified, but without more advanced quantum physics, it is a model that works pretty well.

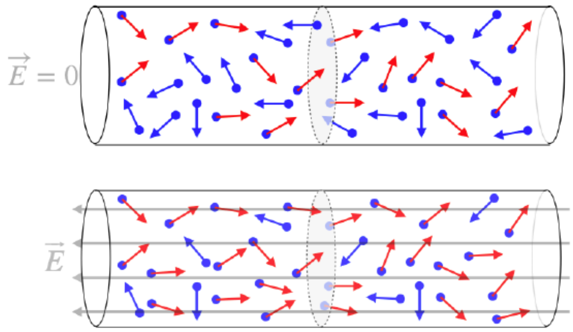
A feature of air resistance that carries over to electrical "drag" is the fact that the faster the object moves, the greater the force. This means that eventually an object moving through the air reaches a terminal velocity, and we will see the same for electrons moving through a conductor. The electric field within the conductor results in a force on the electrons, but the electrons don't keep accelerating indefinitely, just like a falling object under the influence of the gravity force doesn't keep speeding up indefinitely. If we increase the strength of the electric field, then the terminal velocity goes up, just as it would if we increased the gravitational force.

Wait, did we just say the *electric field within the conductor*?! Isn't the electric field inside a conductor always zero? In the case of electro*statics*, yes. That is, we stated previously that the electric field in the presence of a conductor causes charges to migrate, and once they have stopped moving, they produce a second field that cancels the applied field. But now we are discussing what is happening *as the charges move*, so we are now looking at the case when the electric field has not been canceled by the field of a separated charge.

Current Density

We need to take some time to determine the factors that affect the amount of current that passes through a conductor. In a conductor that has no bias placed on it by an electric field, the electrons are still able to move, but they do so randomly, consistent with thermal motion that we studied in Physics 9B. If we watch a specific place in the conductor, we will see these randomly-moving electrons passing by, but the randomness of their motion means that of all the electrons passing the observed checkpoint, half are going each way, for no net charge flow. When an electric field is applied, however, the force it exerts on electrons gives them a bias to move in a specific direction. Of course, some fraction of the electrons will have recently bounced off a nucleus and will briefly be going the opposite direction, but on average the electrons will be flowing in the direction opposite to the electric field.

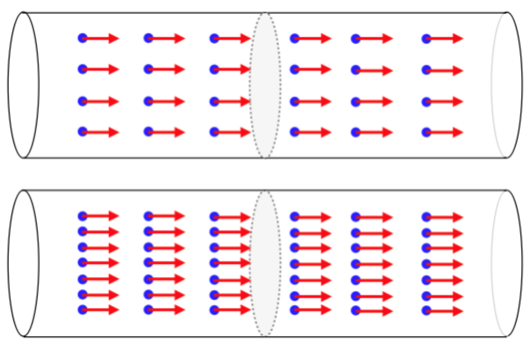
Figure 3.1.1 – Applied Electric Fields Affect Electron Motion in a Conductor



With so many particles crossing a fixed point in so many different ways (and at so many different speeds), we need a way to reconcile our microscopic picture with our simple definition of current above. As we saw in 9B, relating microscopic pictures to macroscopic ones requires speaking in terms of *averages*. In this case, the average we will introduce is called the *drift velocity*, $\overrightarrow{v}_d$. This is a *vector* average – the velocity of the average electron.

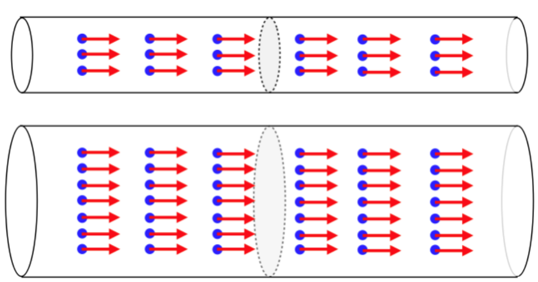
Suppose we know the drift velocity of the electrons – is this enough to give us the current? No, because knowing how fast and in which direction the electrons are moving doesn't include information about *how many* electrons are included in that average, and we need to know the amount of charge passing per second. This number goes up for a fixed drift velocity when the sheer number of electrons goes up. For a given conductor, we can get more electrons moving past a point per second when they are more densely-packed. In the figure below, the electrons in both conductors have equal drift velocities (depicted by the red arrows), but there is more charge passing the checkpoint per second in the lower conductor because the electrons are closer together.

Figure 3.1.2 – Electron Density Affects Current



There is one other consideration to take into account here: Who says that all conductors are the same thickness? The cross-sectional area of the conductor plays a role in the number of charges that can pass through the checkpoint per second.

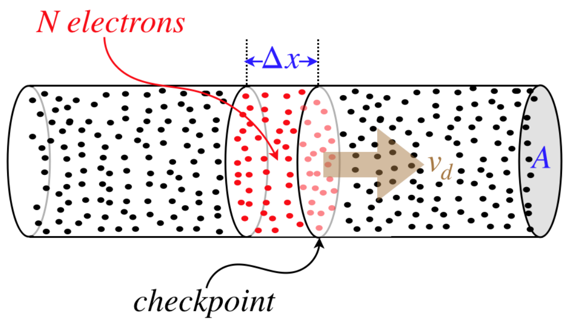
Figure 3.1.3 – Cross-Sectional Area of the Conductor Affects Current



Notice that in the figure above the electrons have the same drift velocity *and* they are equally dense in both conductors, but there is more space for the electrons to pass through the lower conductor, so more charge goes past the checkpoint per second in the lower conductor.

We can put all of these quantities together to come up with a mathematical expression for electric current. If we consider a thin segment of the conductor, as in the figure below, then we can write the rate of charge flow across the checkpoint as the amount of charge in the segment divided by the time it takes for all that charge to exit the segment.

Figure 3.1.4 – Calculating Current



The amount of charge that passes through the checkpoint in the allotted time is Nq, where N is the number of electrons in the tiny volume, and q is the charge of a single electron. So the rate of charge flow is this number divided by the time:

$$I=\frac{Nq}{\Delta t} \tag{3.1.2}$$

The number of particles is equal to the particle density (particles per unit volume, which we will call n), multiplied by the volume of the slice, which is $A\Delta x$. The length of the slice divided by the time that the last electron exits the slice is the drift velocity of the charges, so we get:

$$I=\frac{(nA\Delta x)\,q}{\Delta t}=nqv_dA \tag{3.1.3}$$

Instead of particle density, it is generally more convenient to use our old friend charge density (the three-dimensional variety, ρ), and this is simply the density of particles n multiplied by the charge per particle q, giving:

$$I=\rho v_dA \tag{3.1.4}$$

Example 3.1.1

A thin plastic circular ring is uniformly charged with a total charge of Q. The ring rotates with a rotational speed ω. Find the electric current associated with this charge motion.

Solution

This is not a three-dimensional distribution of charge, so determining the current requires more though than just plugging into what we have found above. The current is the rate at which charge is passing a specific position of the loop (say 12 o'clock). A small slice of the loop has a charge dq on it, and has an arclength we will call ds. These are related to each other through the charge density, in the usual way:

$$dq=\lambda\;ds$$

The charge density is uniform, which means that it equals the total charge divided by the full length over which it is distributed (the circumference of the loop). Calling the radius of the loop R gives us:

$$dq=\frac{Q}{2\pi R}ds$$

Dividing this by the small time it takes the charge to clear that tiny segment gives us the current:

$$I=\frac{dq}{dt}=\frac{Q}{2\pi R}\frac{ds}{dt}$$

The quantity $\frac{ds}{dt}$ is the linear speed of the charge, and we can relate this to the rotational speed, to give our final answer:

$$I = \frac{Q}{2\pi R} v = \frac{Q}{2\pi R}(R\omega) = \frac{Q\omega}{2\pi}$$

While it seems reasonable that the drift velocity of the electrons would be parallel to the axis of a straight conductor like the one in the diagram, in a more general case (such as when the conductor gets wider or thinner), at some positions in the stream the drift velocities could vary from one position to the next. In this case, the definition of "current" will depend upon the area we use as a checkpoint. If the drift velocity at the checkpoint cross section is not perpendicular to the area, then only the component of the drift velocity that is perpendicular will contribute to the current.

With the possibility of different drift velocities at different positions in the stream, we clearly need to add up (i.e. integrate) all of the contributions through a given area. This sounds exactly like the concept of flux we discussed in Section 1.6, except this time the vectors are not electric field vectors. If we pull the area out of Equation 3.1.3, and allow for different drift velocities at different positions (so that we have to integrate just the parts perpendicular to the surface), we get:

$$I = \int \overrightarrow{J} \cdot d\overrightarrow{A}\ , \quad where: \quad \overrightarrow{J}(\overrightarrow{r}) \equiv \rho\, v_d(\overrightarrow{r}) \tag{3.1.5}$$

$\overrightarrow{J}(\overrightarrow{r})$ is called the *current density* (at position $\overrightarrow{r}$).

Alert

Current density is a vector, but the current is not. That is, we define current simply as the rate that charge passes a certain point, and if the flow changes direction (such as in a bend of a wire), the current doesn't change, since it does not have a direction.

Charge Conservation

Consider the flow of charge out of a closed volume. The rate of this flow is related to the total flux of current density out of that volume:

$$-\frac{dQ}{dt} = \oint \overrightarrow{J} \cdot d\overrightarrow{A} \tag{3.1.6}$$

The minus sign appears because the charge within the volume goes down when the current density points out of the volume (in the same direction of the differential area vector). The charge within the volume is the integral of the charge density over the volume, as usual, so:

$$Q = \int \rho dV \quad \Rightarrow \quad -\frac{dQ}{dt} = -\frac{d}{dt}\int \rho dV = -\int \frac{d\rho}{dt} dV \tag{3.1.7}$$

Setting these last two equations equal and using the divergence theorem gives:

$$-\int \frac{d\rho}{dt} dV = \oint \overrightarrow{J} \cdot d\overrightarrow{A} \quad \Rightarrow \quad -\int \frac{d\rho}{dt} dV = \int \left(\overrightarrow{\nabla} \cdot \overrightarrow{J}\right) dV \quad \Rightarrow \quad \overrightarrow{\nabla} \cdot \overrightarrow{J} + \frac{d\rho}{dt} = 0 \tag{3.1.8}$$

This is known as the *continuity equation*. It is the differential statement of what we assumed at the outset – that the rate of charge flow into a closed volume, minus the rate of charge flow out (net flux of current density out), equals the rate at which charge accumulates inside (rate of change of enclosed charge density). Put more simply, it is a differential declaration that charge is neither created nor destroyed – it is conserved. This relation is actually used in many other fields of study (such as fluid mechanics, which we briefly encountered in Physics 9B), where the current is a different kind of flow than that of electric charge. Conservation principles are ubiquitous in physics, and wherever a conservation principle applies, this equation makes an appearance.

3.2: Resistance and Energy Dissipation

Resistivity

We said in the previous section that electric charge flow is caused by the applied electric field, and that what happens to the charges similar to air resistance, inasmuch as a “terminal velocity” is reached (in the case of electricity, this is the drift velocity). We know that increasing the electric field increases the drift velocity and with it the current density, but the exact relationship is not obvious. With experimentation (and/or theoretical models beyond the scope of this course), we find that in fact the electric field vector and the current density vector are directly proportional:

$$\overrightarrow{E} = (constant)\ \overrightarrow{J} \tag{3.2.1}$$

it makes sense that these two vectors would point in the same direction, since the current is defined as the direction of positive charge flow. The constant of proportionality is called the *resistivity*, and is represented by the Greek letter ρ:

$$\overrightarrow{E} = \rho\ \overrightarrow{J} \tag{3.2.2}$$

This relation is known as *Ohm's law*.

Alert

It is common in physics to occasionally see a collision of the same variable used for more than one quantity, such as T for period and temperature, or V for volume and electrostatic potential. But the collision of the use of ρ in Physics 9C is perhaps the most annoying. This letter appears in multiple equations involving current density – both in the drift velocity definition or continuity equation, where it is the charge density, and Ohm's law. Physicists are not bothered by this because they keep the context of equations fresh in their minds, but students encountering these equations for the first time – especially so close together – can find it daunting. The solution is to learn to think of equations in context, keeping a physical system in mind, rather than thinking of them as a jumble of incomprehensible variables.

A quick look at the equation for Ohm's law shows that the greater the resistivity, the less that the moving charges react to the electric field. So what physical properties play a role in this "friction" effect? There are essentially two things that play a role in resistivity, both of them related to the material through which the charge is flowing:

- its molecular structure
- its temperature

The molecular structure comes into play in ways that are impossible to describe in detail without a background in quantum physics. But there are two main ways that the specifics of the type of material involved comes into play. The first is how many free electrons the material allows – all else being equal, the resistivity is lower when more free charge is available. Conductors provide lots of free electrons, semiconductors far fewer, and insulators essentially none. Among conductors, it turns out that the more "regular" (or maybe the word "predictable" is more descriptive) the lattice of fixed nuclei is, the better it conducts (i.e. the lower the resistivity). Metals that are alloys (mixes of different elements) have their atomic nuclei arranged more haphazardly, and therefore have higher resistivities.

The effect of temperature on resistivity for a given material is more intuitive. Recall that what slows the electrons is collisions with atomic nuclei in the lattice of the material. These atoms are constantly vibrating, the energy of which is determined by the temperature of the material. When the temperature rises in a conductor, the more violent vibrations of these obstacles results in more collisions (again, the lattice structure is less "predictable"). So for conductors, the resistivity goes up as the temperature rises. Interestingly, for semiconductors the opposite is true – an increase in temperature results in a lower resistivity. The reason for this is that the resistivity in semiconductors is mostly due to the low number of free electrons. When the temperature is raised, many of the bound electrons gain energy, and tear themselves free. So while the vibrating atoms in the lattice still create more collisions in a semiconductor, the increase in the number of available free electrons is a far more important factor, and the resisitivity goes down.

We can approximate the reaction of resistivity to temperature with a linear relationship. If we measure the resistivity of a material at temperature T_o to be ρ_o, then the resistivity at a new temperature T is given by:

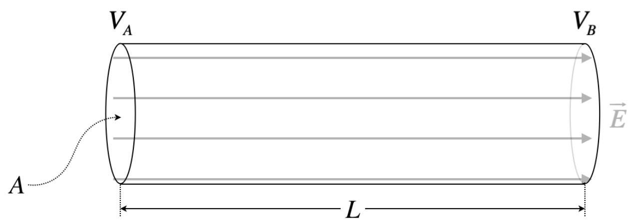
$$\rho(T) = \rho_o\left[1+\alpha\left(T-T_o\right)\right] \tag{3.2.3}$$

The constant α is called the *coefficient of resistivity*. We are not unfamiliar with linear approximations of physical properties like this one. For example, we saw something very similar (it also involved temperature!) in the case of thermal expansion in Physics 9B.

Resistance

As is usually the case, it is easier to deal with scalar energies than with vector forces, so we seek a way to get away from our use of electric field and current density. To discuss energy, we need to shift over to using electrostatic potential rather than electric field. Consider the simple physical system that consists of a cylindrical conductor with a cross-sectional area A, a length L, and a potential difference $V_A - V_B$ (recall that we no longer consider conductors to be equipotentials):

Figure 3.2.1 – Conducting Cylinder



We can now relate the potential difference to the constant electric field, which we can then relate to the current density via Ohm's law, and the current density can then be related to the electric current:

$$\left.\begin{array}{lll} V_A - V_B = \int\limits_A^B \vec{E}\cdot\vec{dl} = E\cdot L & \Rightarrow & E = \dfrac{V}{L} \\ E = \rho\, J & \Rightarrow & E = \rho\dfrac{I}{A} \end{array}\right\} \quad V = I\left(\frac{\rho L}{A}\right) = IR \tag{3.2.4}$$

The quantity $R \equiv \frac{\rho L}{A}$ contains all the information about the conductor needed to determine the current from the voltage difference: its length, cross-sectional area, and resistivity. This value R is called the *resistance* of the conductor. Its units can be easily determined from the equation above – they are volts per amp. This unit is given is own name – *ohms* (Ω). This equation that relates voltage drop, current, and resistance is just a simplified version of Ohm's law, and is usually referred to by the same name (we will do so henceforth).

Power

From this energy perspective, we can see that the charge drops in potential energy when it goes from the higher potential to the lower (okay, technically, it is the negatively-charged electrons that go from lower potential to higher, but that is still a decrease in potential energy). But we also know that the drift velocity doesn't change, so the lost potential energy doesn't go into kinetic energy. Where does it go? Like air friction, electrical resistance results in energy being converted to thermal energy. This means that the conductor with resistance will get hotter as current flows through it.

As we are now talking about flowing charge, it is easier to talk about the *rate* at which energy is converted from electrical potential energy to thermal energy. We know that when a charge q drops through the potential V, it loses a potential energy equal to $U = qV$. The rate at which this occurs (i.e. the power) is the time rate of change of this. Since the voltage remains fixed, we have:

$$P = \frac{d}{dt}(qV) = \frac{dq}{dt}V = IV \tag{3.2.5}$$

We can use Ohm's law to express this relation in two other ways:

$$P = IV = I^2R = \frac{V^2}{R} \tag{3.2.6}$$

This set of equations for power bear a striking resemblance to the set of equations for potential energy in a capacitor given in Equations 2.4.11 (with the exception of the absence of a factor of $\frac{1}{2}$ everywhere). Like those equations, the choice of which of these expressions to use will often depend upon what quantities are unchanging in the physical situation. For example, if you increase the resistance and put the same voltage across it, you find that the rate of energy conversion to thermal is *lower*. But if instead the current is held fixed, the rate of energy conversion to thermal is increased with an increase in resistance.

Alert

One of the most common examples we will use in discussing power in electrical circuits is the light bulb, as it provides a nice visual manifestation of the conversion of electrical energy to another form that exits the circuit. Using our formulas for power, it is clear that the power ceases to be converted the moment that the current stops flowing, but one might notice that (for example) an incandescent light bulb continues to glow a short time after the connection is broken. This does not mean that the current continues flowing for a short time afterward – as far as we are concerned, the current ceases essentially immediately – the continued glow comes from the fact that the glow originates from the increased temperature of the filament, and cutting the current then allows the filament to cool at whatever rate is natural for it to cool in its surroundings, and the (visible) light goes out only when it gets below a certain temperature. So this lag in the cooling time of the filament is responsible for the light going out slowly, it is not a lag in the diminishment of the current.

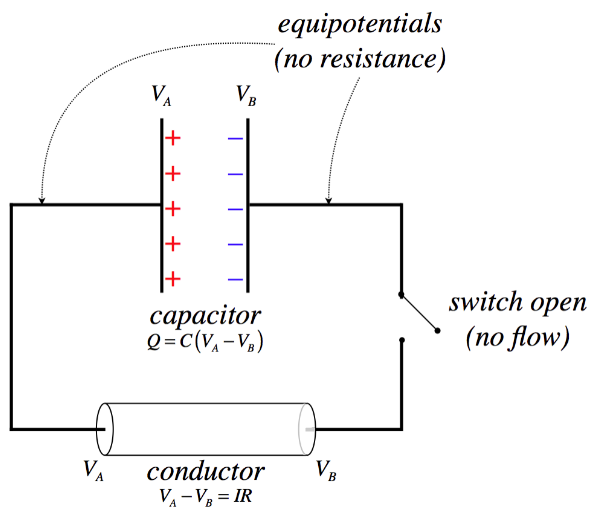
3.3: Networks of Batteries and Resistors

Electromotive Force

We know that voltage differences drive electric currents through resistive materials, but where do these voltage differences come from? Up to now (with capacitors), we said that a voltage difference comes from a separated charge, but to separate that charge we need to move it, and if we do that with a voltage difference, we get into a vicious loop. We obviously need some influence other than a voltage difference due to separated charge in order to separate the charge in the first place. It turns out there are many such mechanisms, but we lump them all in to a general category known as *electromotive force*, or just *emf*. There is only one source of emf that we will look at in detail, and that won't be for another couple chapters. For now, we are going to treat it as a black box, and call it a battery. Electromotive force has the effect of creating a voltage difference, so it has units of volts, but we will label it differently than say a voltage difference across a capacitor. We will use $\mathcal{E}$.

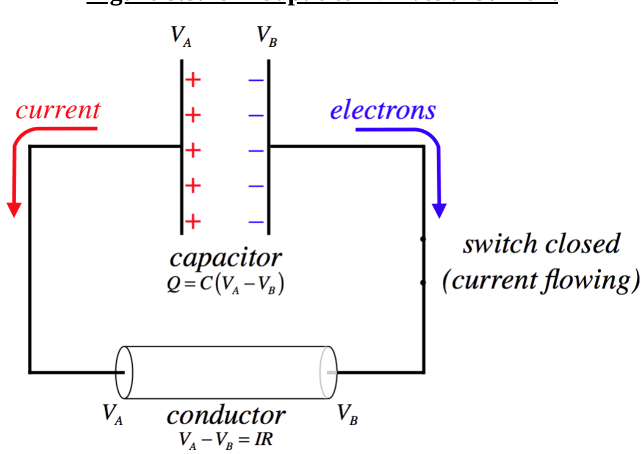
The concept of emf is required primarily to explain what is going on in something called a *circuit*. A circuit is essentially a closed loop, where moving charge gets recycled. Let's look at it from the perspective of a capacitor.

Figure 3.3.1a – Capacitor Drives a Current



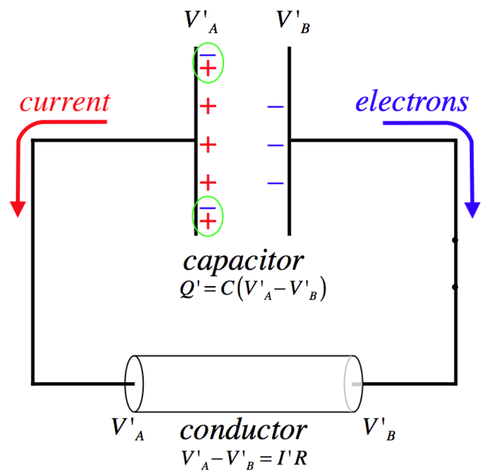
This figure is an abstraction of an actual circuit. In an actual circuit, there is a capacitor and some wires, along with a switch. Here we have collected all the resistive properties into a cylinder that we are calling the conductor. The lines that connect this to the capacitor are equipotentials (*not wires!*) that are introduced as a means for displaying the circuit. In any case, the separated charge on the capacitor comes with a potential difference across the plates, which then also (when the switch is closed) will become the same potential difference across the conductor.

Figure 3.3.1b – Capacitor Drives a Current



When the switch is closed, the potential difference due to the separated charge creates an electric field through the conductor, which exerts a force on the electrons, causing them to flow clockwise. Electric current is defined as the flow of positive charge, so it is defined to be in the opposite direction. The rate of flow is determined by the potential difference and the resistance. The resistance remains relatively constant, but what about the voltage difference?

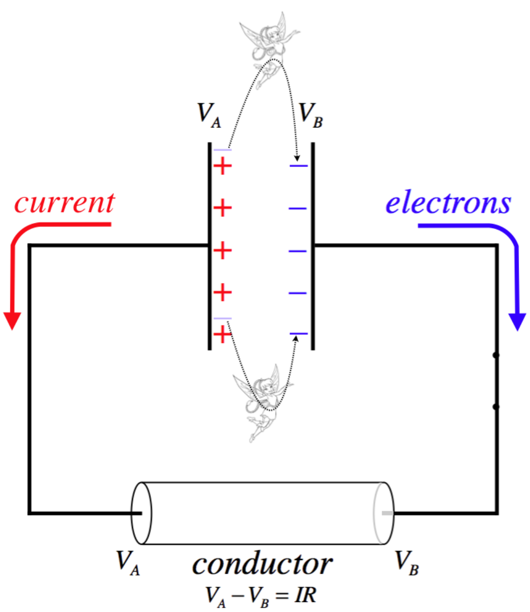
Figure 3.3.1c – Capacitor Drives a Current



When electrons reach their destination on the positively-charged plate, they cancel some of the positive charge, thereby reducing the total charge on the capacitor. The capacitance doesn't change, so less charge corresponds to a smaller potential difference. A reduced potential difference yield a lower current through the conductor.

We are interested in maintaining a *steady* current through a conductor, and our definition of emf is such that it is that which *maintains* the potential difference so that the current can remain steady. So for example, if there existed magical "emf fairies" who, whenever an electron arrived on the positive plate, grabbed it and carried back to the negative plate, then they would be a source of emf.

Figure 3.3.2 – Electromotive Force Drives "Uphill" Charge Transportation



Naturally the transportation of charge back to where it started results in an increase of electrical potential energy in the system (to make up for what was converted to thermal when the current passed through the resistance), so the emf source must be drawing energy from elsewhere. In most real-world applications, we don't rely upon this energy source being the fairy world.

Standard batteries convert chemical potential energy (released through chemical reactions) to drive this charge transportation. As stated earlier, we will see another means for doing this in a future chapter.

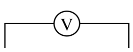
More Symbolic Diagrams

Now that we are discussing current flow, we need to add some more symbols to the collection we started in Section 2.6. In order to accommodate current flow, we now need a way to express resistance in symbolic form. This poses a bit of a problem, as resistance is ubiquitous – conducting wires that connect components to each other in a circuit, and the internal conducting parts contained in batteries, both exhibit resistance, for example.

resistance: 

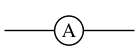
There are electrical components called *resistors* whose sole purpose is to provide resistance to part of a circuit, but use of this symbol goes beyond that single application. For example, if one wants to incorporate the resistance present in a wire in a symbolic diagram, they will use straight lines (equipotentials) to specify where that wire is connected, and will also include one of these resistance symbols to indicate that this segment of wire comes with some resistance. The same is true with the case of *internal resistance* in a component like a battery – if the resistance in the battery is important, the battery is represented in the symbolic diagram in two parts: The usual battery symbol to cover the emf it provides, and a separate resistance symbol to account for the battery's internal resistance. Very often the resistance in a wire or a battery is negligible compared to components in the circuit, and in these cases it's okay to approximate the resistance of the wire as zero, and representing wires with straight lines is acceptable.

In our discussion of static networks of capacitors, we didn't talk about *how* quantities like charge and potential difference are measured. While we won't go into the inner workings of measuring devices used in electrical circuits, it will be useful to introduce symbols for them.

voltmeter: 

This device measures the voltage difference between the equipotentials on either side of it. This device can be used as a "external probe" by connecting the two equipotentials protruding from it to any two places in a circuit, and the meter will show the voltage difference between these two points. This probe is "external" because it can be used without disturbing the circuit in any way.

This meter will also provide the information regarding which of the two equipotentials is at the higher potential. So for example, if one connects a voltmeter leads to the opposite sides of a battery, then it will read the voltage of the battery and which side of the battery is at the higher potential. This device has what is effectively an infinite resistance, so that the circuit can be probed without affecting what is going on in the circuit, since none of the current flow will flow through the voltmeter when it is attached.

ammeter: 

This device measures the amount of current that flows through it, including the direction in which the current is flowing. Unlike the voltmeter, this device cannot be connected to two points in a circuit as an external probe, but rather functions as an "internal probe." In order to measure the current through a component, one of the wires connecting that component must be disconnected from the circuit, and the ammeter inserted between the component and the rest of the circuit where there was previously only the wire. In order to not affect what is actually happening in the circuit, this device must essentially behave like an equipotential, which means it must have effectively zero resistance.

Systems of Resistors

When we combined capacitors in a circuit, we found a way to put them together into equivalent capacitances depending upon whether they were connected in series or parallel. We can do the same with resistors. The definitions of parallel and series are the same as before, because the same principles of equipotentials and charge conservation still hold.

Figure 3.3.3 – Resistors in Series

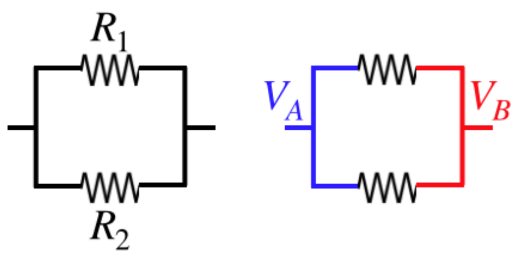
The voltage drop across resistor R_1 is $V_A - V_B$, and across R_2 is $V_B - V_C$ so as with the case of the series capacitors, the sum of the individual voltage drops equals the total voltage drop across both resistors combined:

$$V_{tot} = V_1 + V_2 \tag{3.3.1}$$

Also, we insist that no net charge builds up within, or is lost from, the resistors, so whatever current enters resistor R_1 must also exit it, and then enter resistor R_2 (and then exit it), so the current through the combination equals the current through each individual component:

$$I_{tot} = I_1 = I_2 \tag{3.3.2}$$

Figure 3.3.4 – Resistors in Parallel



The voltage drops across the resistors are the same:

$$V_1 = V_A - V_B = V_2 \tag{3.3.3}$$

Whatever current is entering one side of the combination must leave the right side, and must be divided between the two branches it can follow:

$$I_{tot} = I_1 + I + 2 \tag{3.3.4}$$

Given that all of the rules for voltage drops and disposition of charge are the same for capacitors and resistors, we can jump straight to the answers for equivalent resistance. To do this, however, we have to note that Ohm's law is slightly different from the definition of capacitance. If we make the association of charge with current, we get an inverse parallel between capacitance and resistance:

$$\left.\begin{array}{l} V = Q\dfrac{1}{C} \\ V = IR \end{array}\right\} \quad \Rightarrow \quad \frac{1}{C} \leftrightarrow R \tag{3.3.5}$$

So we can recycle the equivalence equations found for capacitance if we just replace C with $\dfrac{1}{R}$:

$$\left.\begin{array}{l} V_{tot} = IR_{eq} \\ V_1 = IR_1 \\ V_2 = IR_2 \\ V_{tot} = V_1 + V_2 \end{array}\right\} \quad R_{eq} = R_1 + R_2 \quad \text{(series)} \tag{3.3.6}$$

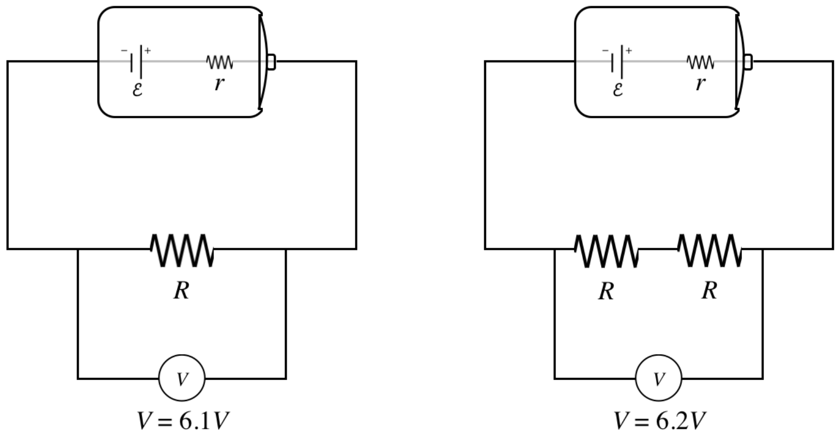
$$\left.\begin{array}{l} I_{tot} = \dfrac{V}{R_{eq}} \\ I_1 = \dfrac{V}{R_1} \\ I_2 = \dfrac{V}{R_2} \\ I_{tot} = I_1 + I_2 \end{array}\right\} \quad \frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} \quad \text{(parallel)} \tag{3.3.7}$$

Internal Resistance

Consider the following puzzle: If we connect a single resistor to a real battery and measure the voltage drop across it with a voltmeter we find it to be $6.1V$. Then we double the number of resistors, and repeat the voltmeter measurement, and much to our surprise, find that the voltage drop is a little bit higher: $6.2V$. How can this be?

Real batteries contain real conductors, and therefore exhibit their own bit of resistance. We refer to this as the *internal resistance* of the battery, and the resistance outside the battery is known as the *load*.

Figure 3.3.5 – Effect of Load on a Real Battery



The math works out like this… First find the equivalent resistance of each circuit, and use it to determine the current that flows through it:

$$\begin{array}{llll} \text{smaller load:} & R_{eq}=R+r & \Rightarrow & I=\dfrac{\mathcal{E}}{R+r} \\ \text{larger load:} & R_{eq}=2R+r & \Rightarrow & I=\dfrac{\mathcal{E}}{2R+r} \end{array} \tag{3.3.8}$$

So there is a different current flowing through each load, and combining these different currents with the different total resistance yields a slightly different voltage drop for each:

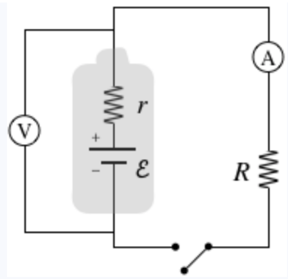
$$\begin{array}{llllllll} \text{smaller load:} & V=IR & = & \left(\dfrac{\mathcal{E}}{R+r}\right)R & = & \left(\dfrac{R}{R+r}\right)\mathcal{E} & = & 6.1V \\ \text{larger load:} & V=I\left(2R\right) & = & \left(\dfrac{\mathcal{E}}{2R+r}\right)\left(2R\right) & = & \left(\dfrac{2R}{2R+r}\right)\mathcal{E} & = & 6.2V \end{array} \tag{3.3.9}$$

The bigger the load is in comparison to the internal resistance, the closer the voltage measured across the load gets to equaling the emf supplied by the battery, because the impact of the r in the denominator on the overall calculation is diminished.

Note that when we draw a diagram, the little symbol is for emf, not for a real battery. If a real battery is intended, then either a battery appears in the picture, or the internal resistance is represented by a symbol for a resistor. The potential difference measured across the two battery leads (or "terminals") is called the *terminal voltage*, and is less than the emf by an amount equal to the voltage drop caused by the internal resistance. The terminal voltage clearly varies according to load, because the amount of current passing through the internal resistance varies according to load and the emf is fixed.

Example 3.3.1

When the switch in the circuit shown in the diagram is open, the voltmeter reads $6.09V$. *When the switch is closed, the voltmeter reading drops to* $5.92V$, *and the ammeter reads* $1.44A$.



a. *Find what the ammeter reads before the switch is closed.*
b. *Find the internal resistance r and emf $\mathcal{E}$ of the battery.*
c. *Find the resistance R of the load.*

Solution

a. With the switch open, there is no potential difference across the load, so there is no current flowing through it. The load is in series with the ammeter, so if there is no current through the resistor, there is also no current through the ammeter.

b. When the switch is open, no current is flowing at all (we assume the voltmeter is ideal, so it has infinite resistance and no current will flow through it), so there is no voltage drop across the internal resistance. Therefore the voltmeter reads the emf of the battery when the switch is open:

$$\mathcal{E} = 6.09V$$

When the circuit is closed, the ammeter reads a current of $1.44A$ passing through the resistor, and since the ammeter is in series with the battery, this is the current flowing through the battery's internal resistance. The potential change measured by the voltmeter in this case is the emf supplied by the battery minus the voltage drop of the internal resistance, so:

$$\mathcal{E} - Ir = \Delta V \quad \Rightarrow \quad r = \frac{\mathcal{E} - \Delta V}{I} = \frac{6.09V - 5.92V}{1.44A} = 0.118\Omega$$

c. The voltmeter also measures the voltage drop across the load, so with the measured current, we get the resistance:

$$R = \frac{\Delta V}{I} = \frac{5.92V}{1.44A} = 4.11\Omega$$

Fuses and Circuit Breakers

In general we prefer not to dissipate too much energy in the internal resistance of a battery. All this does is warm the battery, and the energy that could be used to make a toy function is lost to the environment as thermal energy. As we will see a bit later, the emf we get from our wall sockets is not coming from a battery, and therefore doesn't have an issue with internal resistance. The only systemic resistance present is the resistance in the wires themselves. This turns out to be a double-edged sword.

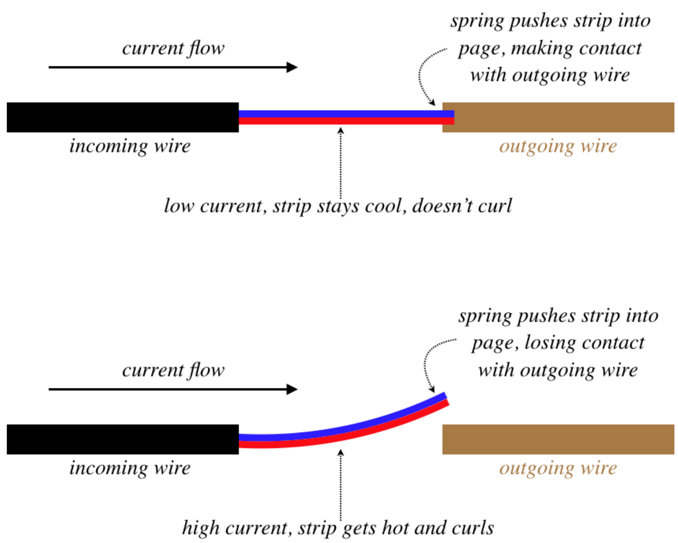
Electrical devices that we plug in are designed to function at a specific voltage, namely the voltage difference provided by our outlets ($110V$). If we want to use two devices at once (say the TV and the DVR), then we can't plug these in series, or they would each get less than 110 volts. We therefore connect them in parallel. Let's suppose we want to plug in many such devices at once (into the same outlet). What happens to the total current drawn from the outlet? Well, the equivalent resistance goes down as we add more devices in parallel, making the total current go up. Now consider what happens to the power dissipated by the wires in the house. The resistance in these wires is small, so when we add enough devices the lower limit on resistance is very small, making the upper limit on current very large. With lots of current passing through the house wires, they get very hot. This is obviously a safety problem. How do we prevent houses from bursting into flames when we plug in too many devices?

The trick is to put a controlled weak link into the wiring. In the old days, this consisted of devices called *fuses*. A fuse consisted of a thin, wispy little wire, enclosed safely within a small module that screwed into a "fuse box" that connected into the wiring somewhere in the house. Because the wire within a fuse was so thin, it would melt when the current got to be too high, thereby breaking the circuit. The amount of current needed to melt this weak link was significantly less than the current needed to cause a fire through the wiring in the walls.

Fuses are still used in individual devices, but you don't really see them in houses anymore. Good thing, too: Home owners who had faulty wiring got tired of constantly buying and replacing fuses, and discovered that they could "solve" the problem by placing a penny (which coincidentally had about the same diameter as fuses) in place of the fuse. Of course, a penny is not a very weak link, and many fires resulted from this dangerous practice.

Nowadays we use devices known as "circuit breakers" in place of fuses. One basic design for a circuit breaker looks like this:

Figure 3.3.6 – Design of a Circuit Breaker



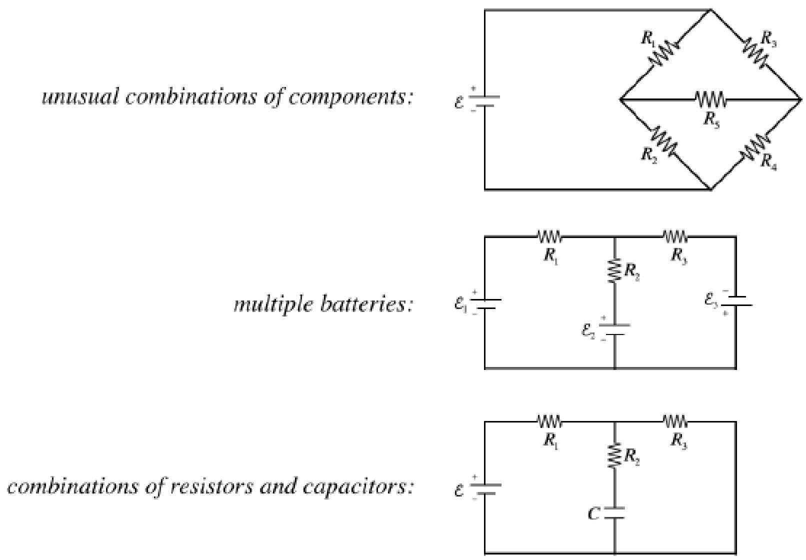
The working parts of this device consist of a bimetallic strip (see the end of Section 5.2 of the Physics 9B LibreText for details on this ingenious device), and a spring. The figure above shows a top view of the circuit breaker, and the spring pushes down on the right end of the bimetallic strip. When the current is low, the temperature of the strip remains low, and it doesn't curl, so the spring pushes the strip into the outgoing wire, making the connection. But when the strip gets too hot, it curls past the outgoing wire, breaking the connection, and the spring pushes the strip so that the connection is not reestablished when the strip cools again (otherwise the electricity through that circuit would surge on-and-off intermittently). Then when the excessive load is removed, the spring is physically compressed, putting the strip back into place. Unlike a fuse, this device can be used over and over.

3.4: Kirchhoff's Rules

Some Problems Cannot Be Solved with R_{eq} and C_{eq}

Despite our ability to reduce circuits using equivalent resistances and capacitances, we can't analyze every circuit imaginable using those shortcuts. For example:

Figure 3.4.1 – Circuits Not Solvable with Equivalent Resistance and Capacitance



It is possible to reduce *fragments* of networks with equivalent resistance/capacitance to simplify our work, but we have to be certain that the elements we are combining truly follow both of the conditions for the type of equivalence we are using. For example, in the third figure above, one might be inclined to proclaim that R_2 and R_3 are in parallel. After all, it's clear that the total current that comes into the junction joining them equals the sum of the currents through each. Well, that's one criterion, but what about the other – that the voltage drops across each is equal? This fails because the capacitor has a voltage drop across it.

Getting Back to Basics

So how do we solve such problems? We do this by using the *same principles* that led to the equivalence formulas, which comes down to two simple rules (called *Kirchhoff's rules*) that are based on charge conservation and energy conservation.

junction rule – Charge remains conserved, so since there is no charge build-up or loss at any junction in a network, the rate at which charge enters a junction equals the rate at which it exits the junction. Put another way, *the current into a junction equals the current out of that junction*.

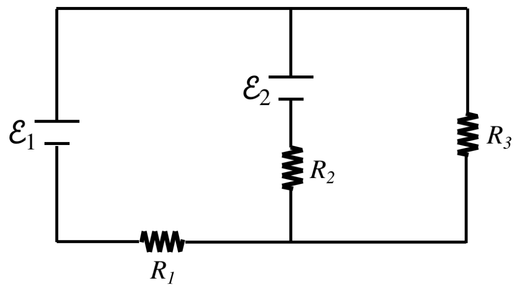
loop rule – Energy remains conserved, which means that when a charge travels around any loop in a network to return to where it started, its potential energy qV should return to the value it had when it was previously at that position. [This of course also presupposes that the emf source is supplying energy to the circuit at the same rate that the circuits resistance is converting energy to thermal, and that the kinetic energy of the charge doesn't change.] Put another way, *the sum of the voltage drops around any closed loop is zero*.

Applying these two rules to simple series and parallel circuits results in the same equivalence rules that we have already, but now these can be used to solve the more complicated problems mentioned above.

Problem Solving

Let's consider a problem involving two batteries and multiple loops:

Figure 3.4.2a – Using Kirchhoff's Rules on a Network

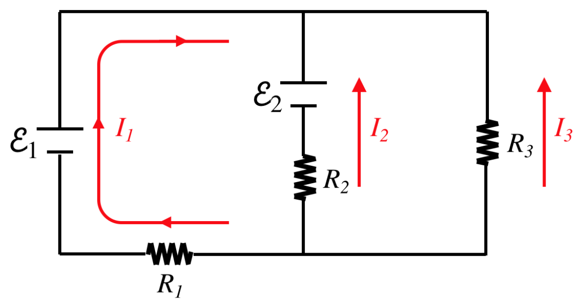


Our goal is to find the current that runs through each of the resistors. There are some very standard steps to follow, and the most important thing to remember is that there exist many choices for these steps, and *none of these choices is wrong*. Don't get slowed down by trying to decide on the "correct" choice to make – every choice will get to the same answer! Without further ado, the steps:

1. **label the currents**

We are solving for the currents, so we need some variables to solve for. But a variable alone is not enough, we also need to label a *direction* for that current.

Figure 3.4.2b – Using Kirchhoff's Rules on a Network



Wait, these currents can't be right – they are all converging on the same junction! No, this labeling is perfectly fine, because in the end we will solve a system of equations, and one or two of these currents will end up being a negative number, indicating that the actual direction of the current is opposite to what we have labeled. We make these labels to solve the problem, and there is no need to be concerned with guessing the actual direction of current flow.

2. **apply the junction rule**

Identify all of the junctions in the diagram (in this case, there are two). We will not require all of the junctions, as we will see here. Choosing the upper junction in this case, and setting the incoming current equal to the outgoing current gives:

$$current\ in = I_1 + I_2 + I_3 = 0 = current\ out \tag{3.4.1}$$

Note that if we choose the other junction, the current in is zero and the current out is the sum of the three individual currents, giving us the same equation. The number of useful junction equations will be one fewer than the total number of junctions. As stated earlier, we can frequently reduce the need for junction equations by using equivalent resistance wherever possible.

3. **apply the loop rule**

Identify all of the loops in the diagram (in this case, there are three – left, right, and outer). As in the case of junctions, we will not require all of the loops (i.e. after we have enough of them, additional loops provide redundant information). The simple way to know if enough loops have been included is to count the number of unknowns and the number of equations. In this case, we have three unknowns (the three currents), and we already have one equation (the junction equation), so we need to use two different loops to attain enough equations to solve for the currents.

This step, while easily stated, comes with many sub-steps. For each loop that is be used, follow the following procedure:

a. **choose a loop direction** – clockwise or counterclockwise
b. **choose a starting point** – any point on the loop will do
c. **follow the loop in the chosen direction and construct a sum of the voltage drops in that direction**

- *When crossing over a battery (or a capacitor)*:
  - Add a positive value equal to the battery's emf if the loop journey crosses from the negative terminal to the positive terminal, because this is an increase in potential.
  - Add a value equal to the negative of the battery's emf if the loop journey crosses from the positive terminal to the negative terminal, because this is a decrease in potential.
- *When crossing over a resistor*:
  - Add a value equal to $-IR$ (where I is the current labeled and R is the resistor encountered) if the direction of the loop journey matches the direction of the labeled current. This is because current always flows from higher to lower potential.
  - Add a value equal to $+IR$ if the direction of the loop journey is opposite to the direction of the labeled current.

d. **set the sum of voltage drops equal to zero**

For the example at hand, this all looks like this (all three loops are provided here, but only two of the equations are needed):

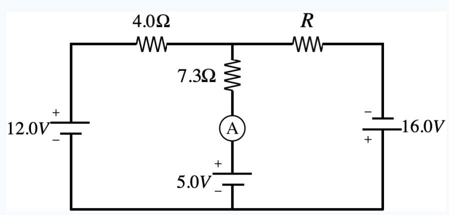
$$\begin{array}{ll} \text{left loop, clockwise, start in lower-left corner:} & +\mathcal{E}_1-\mathcal{E}_2+I_2R_2-I_1R_1=0 \\ \text{right loop, clockwise, start in lower-left corner:} & -I_2R_2+\mathcal{E}_2+I_3R_3=0 \\ \text{outer loop, clockwise, start in lower-left corner:} & +\mathcal{E}_1+I_3R_3-I_1R_1=0 \end{array} \tag{3.4.2}$$

4. **do the algebra** – Solve the simultaneous equations using whatever method you prefer.

Of course, there are many variations on problems – the battery emfs and resistances are not always what is given – but the same principles apply.

Example 3.4.1

Find the resistance R in the network diagrammed below for which the ammeter will measure zero current.



Solution

Noting that there is no current in the central segment and summing the voltage drops clockwise around the left loop (starting at the lower left corner) gives us the current in the outer loop:

$$+12.0V-I\,(4.0\Omega)-(0A)\,(7.3\Omega)-5.0V=0 \quad \Rightarrow \quad I=\frac{7}{4}A$$

Now use that current to sum the voltage drops around the outer loop to find R:

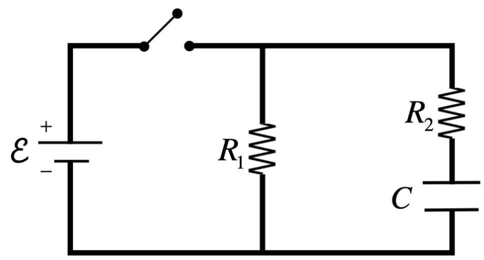
$$+12.0V-I\,(4.0\Omega)-\left(\frac{7}{4}A\right)R+16.0V=0 \quad \Rightarrow \quad R=12.0\Omega$$

3.5: RC Circuits

Capacitors in Networks

We can see how Kirchhoff's rules helps us analyze circuits that either involve awkward combinations of resistors or multiple batteries, but what about including capacitors along with those components? Let's look at a sample of such a network.

Figure 3.5.1 – A Sample Network involving Resistors and Capacitors



The question we want to answer is the usual: "After the switch is closed, what is the current through each resistor?" To answer this question, we start by invoking Kirchhoff's loop rule around the outside loop (clockwise, starting in the lower left corner), and noting that the potential difference between two plates of a capacitor is the ratio of the charge and capacitance, we get:

$$+\mathcal{E}-I_2R_2-\frac{Q}{C}=0 \tag{3.5.1}$$

So in order to ascertain the value of I_2, we need to know how much charge is on the capacitor. Given that charge that flows through the resistor R_2 will be deposited on the plates of the capacitor, it's clear that the amount of charge on the capacitor changes over time. The emf provided by the battery is steady, so this means that the current through the resistor depends upon how much charge started on the capacitor, and how long the switch has been closed. Let's look at two special cases, both of which involve the capacitor being uncharged when the switch is closed.

no initial charge on capacitor, just after the switch is closed

At the moment when the switch is closed, there has not yet been any time for charge to accumulate on the capacitor. With zero charge on it, the voltage difference between the plates is zero. Plugging this into the loop equation above reveals that the current through the resistor is exactly what it would be if the capacitor were not even present. This will of course not remain the case, as the capacitor will begin charging, but at the moment when the current starts, the capacitor can simply be ignored. This result is not special to this particular network – it is generally true, no matter how complicated the network might be:

A capacitor that contains zero charge at an instant in time can be treated as an equipotential within the network at that moment.

capacitor fully charged, a long time after the switch is closed

When the capacitor has been allowed to charge a long time, it will become "full," meaning that the potential difference created by the accrued charge balances the applied potential. In this case, the first and third terms of the Kirchhoff loop equation for the outer loop cancel, which means that no current passes through resistor R_2. In a direct current network, the charge can only accumulate on a capacitor (it doesn't come back off), so it doesn't matter how complicated the network is, given a long enough period of time, the capacitor will fill, and will stop all the current flowing through that branch from flowing.

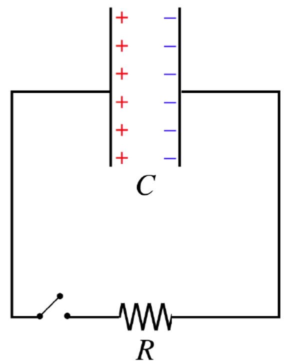
A capacitor that has spent a long time in a closed network will be fully charged, and will not allow any current to pass through the branch it occupies, so it can be treated as if it is an open switch.

You may be wondering how a capacitor (which provides a gap in the conductor) is different from simply a break in the wire. That is, we know that if we cut the wire, the light bulb goes out immediately, while a capacitor allows it to shine (until it is fully charged). The answer is that they are in fact the same! Think about the capacitor that cutting a wire creates: It is the thickness of the wire (very thin), and is typically separated by a very large distance. Using the parallel-plate case as a model, this results in a very small capacitance, which means that for the voltage supplied the amount of charge required to "fully charge" it is very small. The current will not take long at all to do this, so the light shines for only an extremely short time. In the limit where we treat "cutting the wire" as completely disassociating the two open ends, there is zero capacitance, which means zero charge, which means no current through the resistor at all.

A Discharging Capacitor

Now we need to figure out what happens *during* the time period when a capacitor is charging. We start with the most basic case – a capacitor that is discharging by sending its charge through a resistor. We actually mentioned this case back when we first discussed emf. As we said then, the capacitor can drive a current, but as the charge on the capacitor neutralizes itself, the current will diminish.

Figure 3.5.2 – A Discharging Capacitor



We will assume that the capacitor starts with a charge equal to Q_o. We seek to determine everything there is to know about the circuit (charge on the capacitor Q, current through the resistor I, etc.) at a time t if the switch is closed at time $t=0$. Start by using Kirchhoff's loop rule to relate the voltage differences across the two components at some arbitrary time t. Let's label the current so that it is going in the direction we know it must go (counterclockwise), and choose our loop direction the same way. Then starting in the upper-right corner, we see that when we jump across the capacitor we see an increase in potential equal to the voltage across the capacitor, and a decrease in potential across the resistor (because the loop direction matches the labeled current direction).

$$+\frac{Q}{C}-IR=0 \quad \Rightarrow \quad I=\frac{1}{RC}Q\left(t\right) \tag{3.5.2}$$

Now we need to relate the current through the resistor to the charge on the capacitor. Clearly the current is the rate at which charge is leaving the capacitor, which means $I=\left|\frac{dQ}{dt}\right|$, but what should the sign be? Well, since $Q\left(t\right)$ is getting *smaller* as the current flows in the direction we selected, it must be that a positive current equals the negative of the rate of change of the charge on the capacitor. Plugging this in gives:

$$-\frac{dQ}{dt}=\frac{1}{RC}Q\left(t\right) \tag{3.5.3}$$

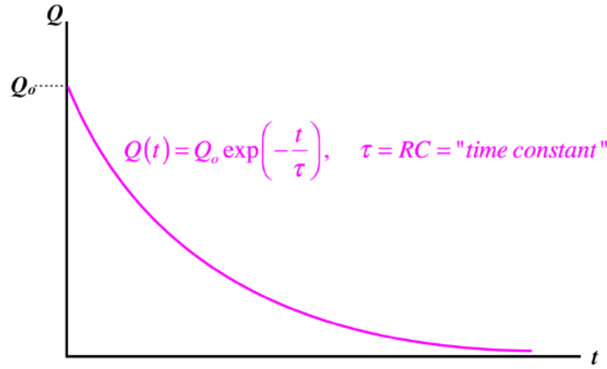
This leaves us with a differential equation that is not difficult to solve. Isolate the variable Q on the left and the variable t on the right, and integrate:

$$-\int\frac{dQ}{Q}=\int\frac{dt}{RC} \quad \Rightarrow \quad -\ln Q=\frac{t}{RC}+const \quad \Rightarrow \quad Q\left(t\right)=Q_oe^{-\frac{t}{RC}} \tag{3.5.4}$$

The final step perhaps requires come clarification. In clearing the natural logarithm function, the additive constant term becomes a multiplicative constant term. To determine what this constant is, we plug in $t=0$ (turning the exponential into a 1) and set $Q\left(t=0\right)$ equal to the charge that the capacitor started with, which we defined to be Q_o.

We therefore find that the charge on the capacitor experiences *exponential decay*. The rate of the decay is governed by the factor of RC in the denominator of the exponential. This value is called the *time constant* of that circuit, and is often designated with the Greek letter τ.

Figure 3.5.3 – Exponential Decay of Charge from Capacitor



Digression: Half-Life

The differential equation that led to the exponential decay behavior for the charge on a capacitor arises in many other areas of physics, such as a fluid transferring through a pipe from one reservoir to another, and nuclear decay. A common way to express the time constant of such a system is in terms of a quantity known as half-life. This is defined as the period of time that must pass for a system to lose half of whatever is decaying (such as the capacitor losing half its charge). It is easy to compute in terms of the time constant:

$$Q\left(t\right)=\frac{1}{2}Q_o=Q_oe^{-\frac{t_{1/2}}{\tau}} \quad\Rightarrow\quad t_{1/2}=\tau\ln 2$$

It is interesting to note that the half-life is the same period of time, no matter what the starting value of the decaying quantity is. So if the charge on a capacitor starts at Q_o and decays for a half-life, then it's new charge is $\frac{Q_o}{2}$, and if it further decays from there for another half-life, then the charge drops to $\frac{Q_o}{4}$.

We can also look at what happens to the current in a discharging capacitor. We already have the relationship between the current and the charge on the capacitor, so it's simple to write down:

$$I\left(t\right)=\frac{1}{RC}Q\left(t\right)=\frac{Q_o}{RC}e^{-\frac{t}{RC}}=I_oe^{-\frac{t}{RC}} \tag{3.5.5}$$

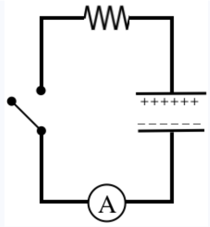
In the final step we used the zero voltage drop around the loop at $t=0$ to replace the combination of constants with the initial current. We see that the current also dies-off exponentially.

One last thing we can look at is the power dissipation. Energy is clearly leaving the capacitor as it charge drops, and energy is leaving the circuit as it is being converted to thermal by the resistor. These rates need to be equal (otherwise, where else is the energy going?), and we see that this is the case:

$$\begin{aligned}\frac{dU}{dt}&=\frac{d}{dt}\left[\frac{Q^2}{2C}\right]=\frac{1}{2C}\frac{d}{dt}\left[Q^2\right]=\frac{1}{2C}\left[2Q\frac{dQ}{dt}\right]=\frac{1}{C}\left[Q_oe^{-\frac{t}{RC}}\right]\left[-\frac{Q_o}{RC}e^{-\frac{t}{RC}}\right]=-\frac{Q_o^2}{RC^2}e^{-2\frac{t}{RC}}\\ P&=-I^2R=\left[I_oe^{-\frac{t}{RC}}\right]^2R=-I_o^2R\,e^{-2\frac{t}{RC}}=-\frac{Q_o^2}{RC^2}e^{-2\frac{t}{RC}}\end{aligned} \tag{3.5.6}$$

Example 3.5.1

The parallel-plate capacitor in the circuit shown is charged and then the switch is closed. At the instant the switch is closed, the current measured through the ammeter is I_o. After a time of $2.4s$ elapses, the current through the ammeter is measured to be $0.60I_o$, and the switch is opened. A substance with a dielectric constant of 1.5 is then inserted between the plates of the capacitor, and the switch is once again closed and not reopened until the ammeter reads zero current.



a. *Find the period of time that elapses between when the switch is closed the second time and when the ammeter reads a current of* $0.20I$.
b. *At the end, all of the electrical potential energy is gone from the capacitor. Find the fraction of this energy that was converted into thermal energy by the resistor.*

Solution

a. We can calculate the time constant from the period of time that the current takes to drop to 0.6 of its original value:

$$0.60I_o = I_o e^{-\frac{t}{\tau}} \quad \Rightarrow \quad \tau = \frac{-t}{\ln 0.60} = \frac{-2.4s}{\ln 0.60} = 4.7s$$

When the dielectric is inserted, the time constant changes. The time constant is proportional to the capacitance, so since inserting the dielectric increases the capacitance by a factor of 1.5, that is the factor by which the time constant changes as well, giving a new time constant of:

$$\tau = RC \quad \Rightarrow \quad \tau_{new} = R(\kappa C) = \kappa\tau_{old} = (1.5)(4.7s) = 7.0s$$

The current is driven by the potential difference across the capacitor, and this is proportional to the charge on the capacitor, so when the current gets down to 60% of its initial value, that means that the charge on the capacitor has dropped by the same factor. To find the time for the current to drop to $0.20I_o$*, we need to know not only the new time constant, but also the new starting current. We can get this from the new starting voltage, which comes from the new starting charge and capacitance:*

$$V_{o(new)} = \frac{Q_{o(new)}}{C_{new}} \quad \Rightarrow \quad I_{o(new)} \frac{V_{o(new)}}{R} = \frac{Q_{o(new)}}{RC_{new}} = \frac{0.60Q_o}{R(1.5C)} = 0.40I_o$$

With the new starting current equal to $0.4I_o$*, we are looking for the time it takes to get down to* $0.2I_o$*, so:*

$$0.20I_o = 0.40I_o e^{\frac{t}{\tau}} \quad \Rightarrow \quad t = \tau \ln 2 = (7.0s)\ln 2 = 4.8s$$

b. We already determined that in the first stage of this process, the charge on the capacitor went down to 60% of its initial amount. This allows us to calculate the energy lost by the capacitor, which is what is converted to thermal:

$$\Delta U_1 = U_o - \frac{(0.60Q_o)^2}{2C} = U_o - 0.36\frac{Q_o^2}{2C} = 0.64U_o$$

So 64% of the energy on the capacitor is converted to thermal energy in the first stage. In the second stage, all of the internal energy in the capacitor is converted, but this amount of energy must be calculated in terms of the new capacitance:

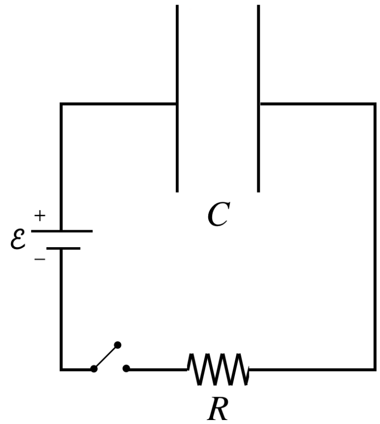
$$\Delta U_2 = \frac{(0.60Q_o)^2}{2(1.5C)} = 0.24U_o$$

So of the original energy stored in the capacitor, 88% of the energy is converted to thermal. Where is the remaining 12%, if all of it is now gone from the capacitor? The field of the capacitor did work drawing the dielectric into it.

A Charging Capacitor

The case of a charging capacitor is not much different, though there are a few nuances to look at. We follow the same procedure as above, starting with the Kirchhoff loop.

Figure 3.5.4 – Charging Capacitor, Initially Uncharged



This time there is a battery included, and the positive lead of the battery charges the positive plate of the capacitor, so following the loop clockwise, with the current defined tin the same direction, and starting in the lower-left corner, results in an increase in potential across the battery, a decrease across the capacitor (goes from positive plate to negative plate), and a decrease across the resistor (in the direction of the current):

$$+\mathcal{E}-\frac{Q}{C}-IR=0 \tag{3.5.7}$$

This time the positive current results in an *increase* to the charge on the capacitor, so the current is related to the charge as the positive derivative, giving us the differential equation:

$$+\mathcal{E}-\frac{Q}{C}-\frac{dQ}{dt}R=0 \quad \Rightarrow \quad \frac{dQ}{dt}=\frac{\mathcal{E}}{R}-\frac{Q}{RC} \tag{3.5.8}$$

Isolating the Q and t variables and integrating:

$$\int\frac{dQ}{Q-\mathcal{E}C}=-\int\frac{dt}{RC} \quad \Rightarrow \quad \ln[Q-\mathcal{E}C]=-\frac{t}{RC}+const \tag{3.5.9}$$

As before, we find the integration constant by plugging in $t=0$. This time the starting charge is zero, so:

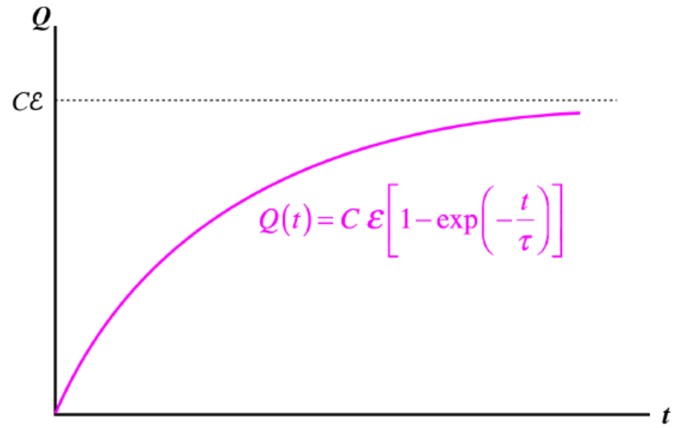
$$const=\ln[-\mathcal{E}C] \quad \Rightarrow \quad -\frac{t}{RC}=\ln[Q-\mathcal{E}C]-\ln[-\mathcal{E}C]=\ln\left[\frac{Q-\mathcal{E}C}{-\mathcal{E}C}\right]=\ln\left[1-\frac{Q}{\mathcal{E}C}\right] \tag{3.5.10}$$

Solving for $Q\left(t\right)$ gives:

$$Q\left(t\right)=C\mathcal{E}\left[1-e^{-\frac{t}{RC}}\right] \tag{3.5.11}$$

So we see the exponential function again making an appearance, but this time it results in the charge asymptotically approaching its maximum (when the capacitor is fully-charged and has a potential across it equal to the battery). The time constant for this is the same as in the discharging case.

Figure 3.5.5 – Charge on Capacitor Asymptotically Approaches a Maximum



The current as a function of time turns out to be identical to that of the discharging capacitor, since the derivative of the constant term in the charging case is zero. That is, the current exponentially decays in both cases, as the system evolves toward equilibrium.

The fate of the energy of this system is a bit more interesting than is was in the obvious case of the discharging capacitor, when all the energy in the capacitor was converted to thermal energy. Here the battery is supplying energy, some of which is lost to thermal energy, and rest is stored in the capacitor. It is interesting to ask how much of the energy goes to each, and the answer is somewhat surprising.

We start by computing the total energy supplied by the battery. We know the *power* supplied by the battery is its voltage multiplied by the current, so we can write down the power supplied as a function of time:

$$P_{battery} = \mathcal{E}I = \mathcal{E}I_o e^{-\frac{t}{RC}} \tag{3.5.12}$$

But we are interested in the total energy supplied over the entire period of time that current is flowing, which means that we need to integrate this power function from $t = 0$ to $t = \infty$:

$$U_{battery} = \mathcal{E}I_o \int_0^\infty e^{-\frac{t}{RC}}\,dt = -\mathcal{E}I_o RC\left[e^{-\frac{t}{RC}}\right]_0^\infty = \mathcal{E}I_o RC \tag{3.5.13}$$

When the switch is first closed, it is as though the capacitor isn't there, so the initial current multiplied by the resistance is the voltage drop across the resister at that moment, which is simply the emf supplied by the battery, giving us:

$$U_{battery} = C\mathcal{E}^2 \tag{3.5.14}$$

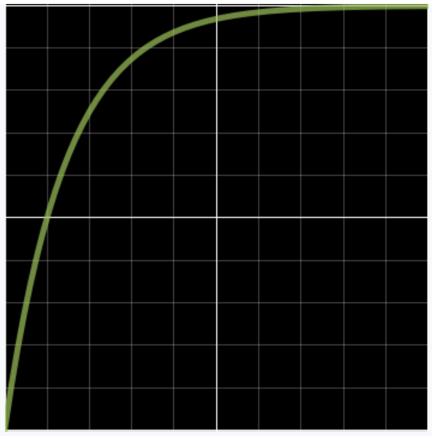
We know that at the end, the capacitor is fully-charged, and therefore is at the same voltage difference as the battery, giving it a total stored energy equal to:

$$U_{capacitor} = \frac{1}{2}C\mathcal{E}^2 \tag{3.5.15}$$

So we see that exactly *half* the energy that comes from the battery goes to the capacitor, which means that the other half is converted to thermal energy by the resistor.

Example 3.5.2

A voltmeter that plots potential differences in real time is connected across the plates of a capacitor as it is charged in a simple circuit that includes the capacitor (which starts with zero charge), a battery, and a resistor all in series. The voltmeter's output is shown below, with each marking along the horizontal axis representing 2 milliseconds and each marking along the vertical axis representing 1 volt. An ohmmeter used to measure the resistance in the circuit gives it to be $1500\ \Omega$



a. *Find emf of the charging battery.*
b. *Find the capacitance of the capacitor.*

Solution

a. The capacitor starts at zero potential difference (it is uncharged), and asymptotically approaches a potential difference of $10V$. *The capacitor stops charging when it reaches the emf of the battery, so the battery's emf is* $10V$.

b. We know the resistance of the circuit, so if we can determine the time constant of the circuit, we can compute the capacitance. This capacitor reaches half its charge after $2\ ms$ *(one horizontal grid line), so this gives us all we need to compute the time constant:*

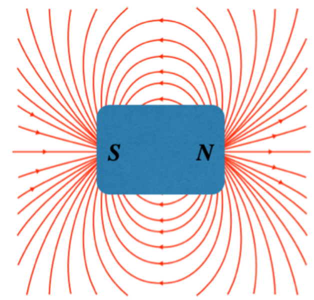
$$\frac{1}{2}Q_o=Q_o\left(1-e^{-\frac{2\ ms}{\tau}}\right)\quad\Rightarrow\quad\tau=\frac{2\ ms}{\ln 2}=2.89ms$$

From this we get the capacitance:

$$\tau=RC\quad\Rightarrow\quad C=\frac{\tau}{R}=\frac{2.89\times10^{-3}s}{1.5\times10^{2}\Omega}=1.9\mu F$$

CHAPTER OVERVIEW

4: MAGNETISM



The phenomena of electricity and magnetism turn out to be two sides of the same coin, but we follow the historical path of examining them as if they are separate forces. We therefore set aside electricity and dive into magnetism.

4.1: MAGNETIC FORCE

We begin our exploration of magnetism with a discussion of the effect that a magnetic field (the source and properties of which we will study later) has on electric charges.

4.2: MAGNETIC MOMENT AND TORQUE

We extend our idea of an electric dipole into magnetism. Given there are no "point charges of magnetism," the idea of a magnetic dipole moment has even more utility than the electric dipole moment has.

4.3: MAGNETIC FIELD

Now that we know how magnetic fields exert forces on moving charges, we need to determine how to compute magnetic fields from their sources. As we did with electric fields, we will not only discuss a general procedure of integrating infinitesimal contributions, but will also catalog the fields for a few common geometries.

4.4: SOURCES OF MAGNETIC FIELDS

Now that we know the basic cause of magnetic fields, we will practice calculating these fields, and will look at some common sources.

4.5: AMPÉRE'S LAW

In electricity, we discussed the useful-if-abstract relationship between charge and field known as Gauss's law. As a close cousin to electricity, it should not be surprising that magnetism has a similar relationship between current and field. As we did with Gauss's law, we will find ways to use this mathematical relation to solve for fields of symmetric arrangements of field sources.

4.1: Magnetic Force

Forces on Moving Charged Particles

If we run currents through two parallel wires, something unexpected happens – the wires exert forces on each other! One might be inclined at first to explain this by claiming that putting currents through the wires puts electric charge into them, and that the electric charges are exerting electrical forces on each other. But this is not correct. Current is simply a steady flow of charge – there isn't any charge built-up in the wire. For every electron that enters the wire, a corresponding electron exits it.

So the force must have something to do with the *motion* of the charges. After further experimentation, we find that if we place a stationary net charge near a conducting wire with current, there is no force between them. So apparently there must be motion for *both* sets of charges in order to exhibit this force.

While this force involves electric charge, it clearly is not electrical in nature. That is, it is altogether different from the Coulomb force. We therefore give it a different name... We call it the *magnetic force*. Like the electric force, we will explain it in terms of a vector field. And as with the electric force, this will require the two-step theory of first explaining how a charge acts as a source for the field, and then how another charge reacts to being in the presence of a field. We are going to hold off on the first step for now, and focus on the second step, which means we will just start with a magnetic field (without worrying about how it got there), and discuss how the field exerts a force on the moving charge.

We will approach this topic as if we were performing experiments to extract the information we want. Here is a list of our observations from these experiments:

- *The strength of the magnetic force on a charge is proportional to the magnetic field through which the charge is moving* – This is not surprising, as it was also true for the electric force and field, but more to the point, it really is a result of our *definition* of magnetic field.

$$\left|\overrightarrow{F}_B\right| \propto \left|\overrightarrow{B}\right| \tag{4.1.1}$$

[*Note: The traditional variable for magnetic field is* B.]

- *The strength of the magnetic force on a charge is proportional to the magnitude of the charge* – Again, not surprising.

$$\left|\overrightarrow{F}_B\right| \propto q\left|\overrightarrow{B}\right| \tag{4.1.2}$$

- *The strength of the magnetic force on a charge is proportional to the speed at which the charge is moving though the field* – This was mentioned above, and it is the first divergence form the electrical case.

$$\left|\overrightarrow{F}_B\right| \propto q\left|\overrightarrow{v}\right|\left|\overrightarrow{B}\right| \tag{4.1.3}$$

- *The strength of the magnetic force on a charge varies depending upon the relative directions of the magnetic field and the charge's velocity vector* – Now *this* is new! Specifically, we find that the force is zero if the charge happens to be moving parallel to the field, and is its strongest when the field and velocity are perpendicular to each other. Further experimentation reveals that the strength of the force is proportional to the sine of the angle between the field and velocity vectors.

$$\left|\overrightarrow{F}_B\right| \propto q\left|\overrightarrow{v}\right|\left|\overrightarrow{B}\right|\sin\theta \tag{4.1.4}$$

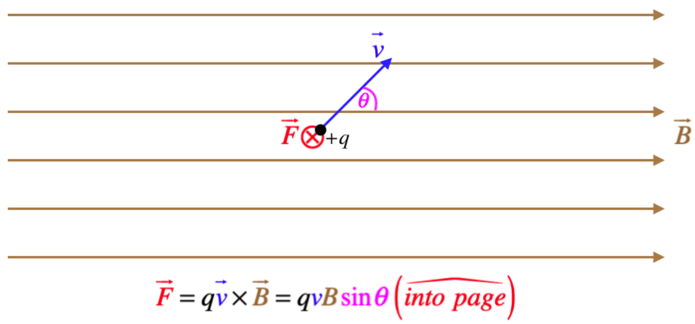
- *The direction of the magnetic force is perpendicular to both the direction of the velocity and the direction of the magnetic field* – This is quite different from the electric case, for which the direction of the force and the field are always in the same or opposite directions.

Perhaps the reader has put together all the puzzle pieces to get to the final expression of the magnetic force on moving point charge:

$$\overrightarrow{F}_B = q\overrightarrow{v} \times \overrightarrow{B} \tag{4.1.5}$$

The figure below shows the relationship of the velocity, field, and force vectors. It should be noted that the symbols $\odot$ and $\otimes$ will start playing common roles in diagrams from here on, and they represent the directions "out of the page" and "into the page," respectively. It is also important to note that like the case of the electric force, a negatively-charged particle experiences a magnetic force in the opposite direction of a positively-charged particle under the same conditions.

Figure 4.1.1 – Magnetic Force Direction from Velocity and Field Directions



[*For a review of the proper use of vector ("cross") products, including the right-hand-rule for determining direction in space, see* *Section 1.2 of the Physics 9A LibreText*.]

There's no reason that electric and magnetic fields can't coexist in the same space. When they do, the force on the point charge is the vector sum of the two forces. The combination (often referred to as the *Lorentz force*) is therefore:

$$\overrightarrow{F} = q\left(\overrightarrow{E} + \overrightarrow{v} \times \overrightarrow{B}\right) \tag{4.1.6}$$

Before moving on, we should say a word about units. For the equation given above to have proper units, the untis for magnetic field must be force divided by charge-times-velocity. Given how many different names we have for various units, we can of course express this many ways, but once again we will give this quantity its own name:

$$[B] = \frac{[F]}{[q]\,[v]} = \frac{N}{C \cdot \frac{m}{s}} = \frac{N}{A \cdot m} \equiv T \quad (\text{" Tesla "}) \tag{4.1.7}$$

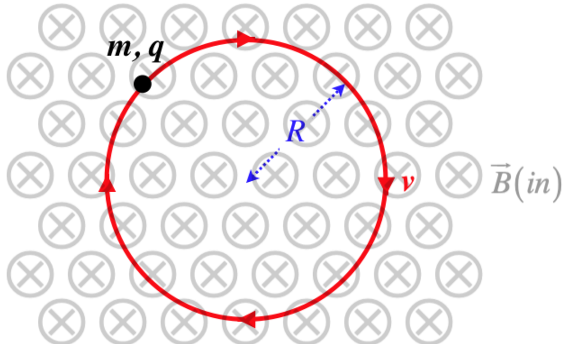
It turns out that a magnetic field with a magnitude of $1T$ is quite strong. It is not beyond common experience (neodymium bar magnets get to this strength), but more commonly encountered magnetic fields (such as that of the Earth, or a compass needle) are significantly less, and it is more common to see these field strengths described in units of *Gauss* (G), which the simply one ten-thousandth of a Tesla.

Charge Motion in a Magnetic Field

A charge within a magnetic field that is subject to no other forces other than the magnetic force, follows a motion with a very interesting property. The magnetic force only acts perpendicular to the direction of motion, which means *the charge can never speed up*. Such a force can only change the direction of motion. This should be clear from basic vector mathematics, but it can perhaps be more easily understood in terms of energy. If the force is always perpendicular to the direction that the particle is moving, then this force can never do any work, which means it can never cause a change in the particle's kinetic energy.

Let's consider a simple case of a charged particle that moves in a plane that is perpendicular to a uniform magnetic field. This charge would experience a force that never changes magnitude (because the charge remains unchanged, the field is uniform, the direction of the velocity is always at right angles to the field, and as we stated above, it can't speed up. This constant-magnitude force is also always perpendicular to the motion of the particle. These two conditions on the force (constant magnitude and always at right-angles to the motion) sound very familiar – these are the conditions for circular motion! Sure enough, such a particle would move in a closed circle. We can use Newton's 2nd law and what we know about centripetal acceleration to put the magnetic force together with the kinematic details of this particle's motion.

Figure 4.1.2 – Charge Moving in a Plane Perpendicular to a Magnetic Field



[*Note: Is the charge in the figure above positive or negative?*]

Ignoring the sign of q (i.e. treating it as an absolute value) we get:

$$\overrightarrow{F} = m\overrightarrow{a}_c \;\;\Rightarrow\;\; qvB\sin 90^o = m\omega^2 R \;\;\Rightarrow\;\; \omega = \frac{q}{m}B \tag{4.1.8}$$

The angular speed of the particle depends only upon its charge, its mass, and the magnetic field strength.

Okay, so what if the particle is not moving in this plane? Suppose it has a component of velocity into or out of this plane. Well, this component will be parallel to the magnetic field, and therefore will not contribute to the force on the particle. The components of the velocity in the plane perpendicular to the field still have the same effect. The result is that the particle's motion parallel to the field is totally unaffected, while it moves in a circle in the plane perpendicular to the field. That is, the particle follows a *helical path*.

Forces on Current-Carrying Wires

We have a vector equation describing the force on a point charge moving through a field, and now we would like to extend this result to currents flowing through conductors in a magnetic field. To do this, we use a common chain-rule trick to relate current through a short length of wire to a moving charge:

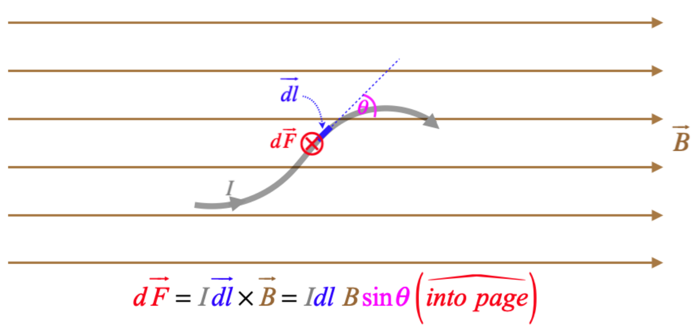
$$I = \frac{dq}{dt} = \frac{dq}{dl}\frac{dl}{dt} \;\;\Rightarrow\;\; I\;dl = dq\frac{dl}{dt} = dq\;v \tag{4.1.9}$$

If we take into account the direction of the current flow through dl and the velocity, we have the vector version of the contribution of a small segment of current, and we can plug it into the force equation:

$$I\;\overrightarrow{dl} = dq\;\overrightarrow{v} \;\;\Rightarrow\;\; d\overrightarrow{F} = dq\;\overrightarrow{v}\times\overrightarrow{B} = I\;\overrightarrow{dl}\times\overrightarrow{B} \tag{4.1.10}$$

Figure 4.1.1 only needs to be altered slightly to depict this situation – the wire may be curved, but a tiny segment of it behaves like the point charge.

Figure 4.1.3 – Force on a Tiny Segment of a Current-Carrying Wire



If the total force on a length of wire is desired, all the infinitesimal contributions to force need to be integrated. This can present a challenge is the wire is not straight, since $\overrightarrow{dl}$ changes direction, or, of course, if the magnetic field changes from one point of the wire to the next. In practice, we frequently deal with straight-line segments in uniform magnetic fields, which yields a simple result for the magnitude (the direction can be determined by the right-hand-rule):

$$\text{force on straight wire of length } L \text{ and current } I \text{ in uniform field } B = ILB \tag{4.1.11}$$

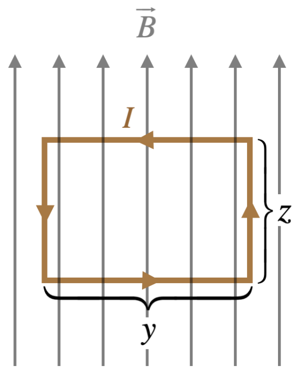
We typically deal with currents in closed circuits, so forces on single segments of wire are of limited use. We move on to this more common case next.

4.2: Magnetic Moment and Torque

Torque on a Loop of Wire

Let's use our result for the force on a segment of wire to analyze the case of the effect of a magnetic field on a closed loop of wire. We will choose a simple geometry for this analysis – a rectangular loop of wire with two sides parallel to a uniform magnetic field.

Figure 4.2.1 – Closed Rectangular Loop of Wire in a Uniform Magnetic Field



Here are the main features of this set-up:

- The vertical sides of the rectangular loop are parallel to the magnetic field, so the force on every element is zero, adding up to a total of zero force on each of those sides.
- The horizontal sides of the rectangular loop are perpendicular to the field, so the sine of the angle that appears in the cross-product is exactly equal to one.
- The magnetic field is uniform, and the current doesn't change direction, so combining this with the previous observation, we get a force on each of these segments equal to the same value: IyB.
- We can work out the directions of the forces on these two segments in two ways: Using the right-hand-rule, or plugging-in the unit vectors for the directions of the current and magnetic field, and computing the cross-product. Doing so reveals that the force on the top segment is into the page, while the force on the bottom segment is out of the page.

The forces on the horizontal segments cancel, resulting in zero net force on the loop, but of course there is a net torque. Choosing an axis that is a horizontal line passing through the centers of the two vertical segments, we can compute the net torque on the loop. Choosing the positive torque direction to be to the left, the forces on top and bottom both generate torques in that positive direction, so:

$$\tau_{net} = F_{top}\left(\frac{z}{2}\right) + F_{bottom}\left(\frac{z}{2}\right) = 2\left(IyB\right)\left(\frac{z}{2}\right) = I\left(yz\right)B \tag{4.2.1}$$

Magnetic Dipole Moment

Here we introduce a shortcut for future torque calculations. The quantity yz is the area of the loop, A. In future applications, we may have the current fed into the loop by a single wire, which is wound around th perimeter several times. The force exerted on each side of the loop (and therefore the torque) will then be multiplied by the number of turns in the wire, N. The product of N, I, and A is written as a single quantity μ, giving the magnitude of the torque for this case the simple form of $\tau = \mu B$.

If this loop turns upon its axis, then the moment arm shrinks. For example, if the top of the loop rotates back and the bottom rotates forward by 90^o, then the forces on those segments will be directly away from each other. These forces act straight through the axis, so the torque they produce is zero. We know that torque and magnetic field are both vectors, and the torque created is related to the orientation of the loop in the field. We can account for the loop orientation by defining a *magnetic dipole moment*:

$$\vec{\mu} \equiv NI\vec{A} \tag{4.2.2}$$

The vector $\vec{A}$ has a magnitude equal to the area of the loop, and has a direction that is perpendicular to the plane of the loop, in the direction defined as follows: Curl the fingers of your right hand in a direction that traces the direction of the current around the loop, and the thumb of that hand points the direction of the vector. For example, the loop in Figure 4.2.1 would have a magnetic moment that points out of the page.

The torque vector can now be calculated from the magnetic dipole moment in the same way that the torque exerted on an electric dipole was calculated:

$$\overrightarrow{\tau}_{electric} = \overrightarrow{p} \times \overrightarrow{E} \quad \Leftrightarrow \quad \overrightarrow{\tau}_{magnetic} = \overrightarrow{\mu} \times \overrightarrow{B} \tag{4.2.3}$$

We can see that this works for the case shown in Figure 4.2.1: The angle between the magnetic dipole moment (which points out of the page) and the magnetic field is 90^o, so the sine of the angle between these vectors that appears in the cross product is 1, giving the answer we found above. When the loop rotates around the horizontal axis, the angle between the magnetic dipole moment and the field changes, reducing the moment arms of the forces by a factor of $\sin\theta$ – exactly the amount accounted-for in the cross-product. When the loop rotates to the point where its plane is perpendicular to the field, the magnetic moment and field are parallel, making the torque zero, as we found above.

Example 4.2.1

A $2.00\ A$ current flows through a circular conductor, which has a radius of $12.0\ cm$ and lies in the x-y plane. When viewed from the + z-axis, the current is flowing clockwise. This loop is in the presence of a uniform magnetic field given by:

$$\overrightarrow{B} = B_o\left(\hat{i} - 3\hat{j} + 2\hat{k}\right)\ , \quad where:\ \ B_o = 1.50T$$

Find the torque (vector) exerted on the conductor.

Solution

To find the torque vector, we first need the magnetic moment. We calculate that to be (use RHR for direction):

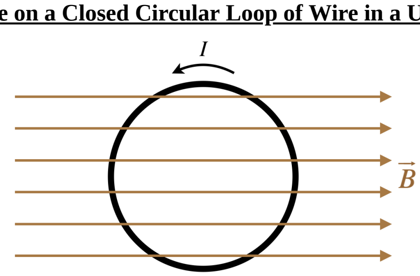
$$\overrightarrow{\mu} = IA\left(-\hat{k}\right) = (2.00A)\,\pi(0.12m)^2\left(-\hat{k}\right) = \left(-9.05\times10^{-2}A\cdot m^2\right)\hat{k}$$

Now just plug into the formula for torque:

$$\overrightarrow{\tau} = \overrightarrow{\mu}\times\overrightarrow{B} = \left[\left(-9.05\times10^{-2}A\cdot m^2\right)\hat{k}\right]\times\left[(1.50T)\left(\hat{i} - 3\hat{j} + 2\hat{k}\right)\right] = -(0.136\ N\cdot m)\left(3\hat{i} + \hat{j}\right)$$

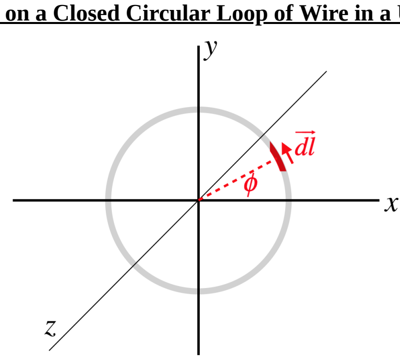
Although we derived the formula for the magnetic dipole moment using a rectangle, it turns out that as long as the loop lies in a plane, the formula works no matter what shape it is. As an illustrative example, we will solve the torque on a circular loop. This is a more difficult example than the rectangle, for reasons that will become clear, but it does demonstrate important tools for integrating infinitesimal contributions and dealing with vector products.

Figure 4.2.2a – Torque on a Closed Circular Loop of Wire in a Uniform Magnetic Field



Much like we did for integrating charge distributions to obtain fields, we start by introducing a coordinate system (*make sure it is right-handed, i.e. choose the axes so that $\hat{i}\times\hat{j} = \hat{k}$*), select an infinitesimal piece os the loop, and describe it in terms of the coordinates, labeling whatever variables we will need to know along the way.

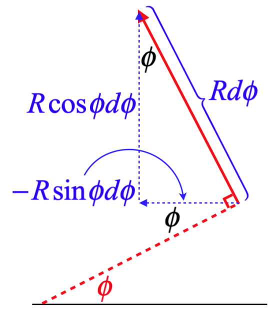
Figure 4.2.2b – Torque on a Closed Circular Loop of Wire in a Uniform Magnetic Field



Here we have chosen to place the loop in the x-y plane, and the magnetic field points in the $+x$-direction. An infinitesimal slice of wire has been selected at an angle ϕ up from the $+x$-axis.

Next we need to express the vector $\overrightarrow{dl}$ mathematically. Its magnitude is the length of an infinitesimal segment of arc, which is $R\,d\phi$. The direction is trickier to come up with, but blowing up the picture and doing a bit of geometry, we can determine its components:

Figure 4.2.3 – Writing the Current Element Vector



Putting it together into a single vector:

$$\overrightarrow{dl} = R\,d\phi\left(-\sin\phi\;\hat{i} + \cos\phi\;\hat{j}\right) \tag{4.2.4}$$

We now have everything we need. As complicated as the geometry is with the force and then the torque, we don't have to track it – all we need to do is do the vector math properly. For example, the force on the current element is:

$$d\overrightarrow{F} = I\overrightarrow{dl}\times\overrightarrow{B} = I\left[R\,d\phi\left(-\sin\phi\;\hat{i} + \cos\phi\;\hat{j}\right)\right]\times\left[B\;\hat{i}\right] \tag{4.2.5}$$

Recalling the cross products of unit vectors from Physics 9A, we plug in $\hat{i}\times\hat{i}=0$ and $\hat{j}\times\hat{i}=-\hat{k}$, and the force on this element becomes:

$$d\overrightarrow{F} = IRB\cos\phi\;d\phi\left(-\hat{k}\right) \tag{4.2.6}$$

To get the torque, we choose the origin as a reference point, and compute the infinitesimal contribution to the torque directly. Plugging in the position vector and doing the vector math gives:

$$d\overrightarrow{\tau} = \overrightarrow{r}\times d\overrightarrow{F} = \left(R\cos\phi\;\hat{i} + R\sin\phi\;\hat{j}\right)\times\left(-IRB\cos\phi\;d\phi\;\hat{k}\right) = -IR^2B\;d\phi\left[\cos^2\phi\left(-\hat{j}\right) + \sin\phi\cos\phi\left(\hat{i}\right)\right] \tag{4.2.7}$$

All that remains is to add up all of the torque contributions, which means integrating over the angle ϕ from $0\rightarrow 2\pi$:

$$\overrightarrow{\tau} = -IR^2B\int_0^{2\pi} d\phi\left[\cos^2\phi\left(-\hat{j}\right) + \sin\phi\cos\phi\left(\hat{i}\right)\right] = IR^2B\left[\pi\;\hat{j} - 0\;\hat{i}\right] = I\left(\pi R^2\right)B\;\hat{j} \tag{4.2.8}$$

Sure enough, the magnitude of the torque comes out to be $\mu\,B$, where $\mu = IA$. And using the right-hand-rule to get the direction of the magnetic moment (out of the page) followed by the direction of the torque from the right-hand-rule applied to $\overrightarrow{\mu}\times\overrightarrow{B}$, confirms that the direction works as well.

This problem seemed very daunting because the direction of $\overrightarrow{dl}$ is changing everywhere on the circle, but once this vector is written in terms of ϕ and the unit vectors, the math does the rest!

Comparing Magnetic to Electric

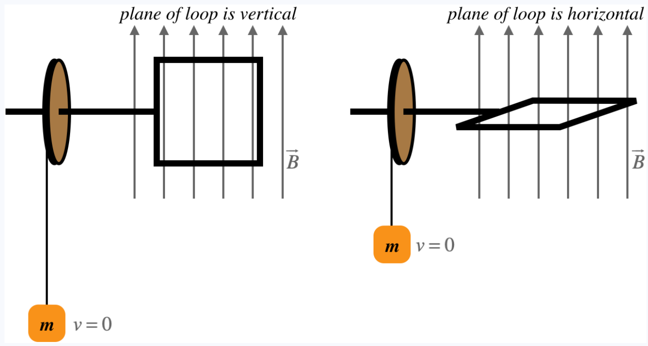
We found that we could do more with electric dipoles than just compute torques, and the same is true for magnetic dipoles. The nice thing is that we don't have to work through everything again – the same results simply translate-over by replacing $\overrightarrow{p}$ with $\overrightarrow{\mu}$, and $\overrightarrow{E}$ with $\overrightarrow{B}$. So we have the potential energy of a dipole in a field:

$$U_{electric} = -\overrightarrow{p}\cdot\overrightarrow{E} \quad\Leftrightarrow\quad U_{magnetic} = -\overrightarrow{\mu}\cdot\overrightarrow{B} \tag{4.2.9}$$

As we saw with the electric dipole, we can explain the instability of the anti-aligned position in terms of having a maximum potential energy.

Example 4.2.2

The device shown in the diagram consists of a square loop through which flows a steady current, located in a uniform magnetic field, and attached to an axle that turns a wheel. Around the wheel is wound a (massless) string (the string comes down over the front of the wheel as shown in the diagram), from which hangs a mass. The mass rises and falls such that it is at rest when the plane of the loop is vertical, and is again at rest when the plane of the loop is horizontal (i.e. the loop does not spin all the way around, but rather goes back-and-forth).



a. *In which direction (as seen by our view of the loop being in the vertical plane) is the current going, clockwise or counter-clockwise?*
b. *Use the values below to determine the amount of current flowing through the loop.*

$$g = 9.8\ \frac{m}{s^2}\ ,\quad m = 0.45\ kg\ ,\quad B = 1.2\ T\ ,\quad \text{length of the sides of the loop} = 0.54m\ ,\quad \text{radius of wheel} = 0.25m$$

Solution

a. The magnetic field turns the loop such that the top segment goes into the page. Curling our fingers in that direction gives a direction of torque that is to the left. The torque direction comes from the cross product of magnetic moment and field. The field is up, so for the cross product to be to the left, the fingers have to curl from out of the page upward. Therefore the magnetic moment points out of the page. Curling our fingers counterclockwise gives this direction, so that is the direction of the current.

b. The system conserves energy, and the mass is moving at neither the top nor the bottom of its journey, so the kinetic energy doesn't change. Therefore all of the energy changes are in the form of potential energy. At the top of its journey the mass gains gravitational potential energy, and this must have come from the magnetic potential energy of the dipole in the field. The wheel turns through an angle of $\frac{\pi}{2}$, so we can figure out how much string is wound up by the wheel (which equals the change in height of the mass). This is then used to find the change in gravitational PE:

$$\Delta y = r\theta = (0.25\ m)\left(\frac{\pi}{2}\right) = 0.39\ m \quad \Rightarrow \quad \Delta U_{grav} = mg\Delta y = (0.45\ kg)\left(9.8\ \frac{m}{s^2}\right)(0.39\ m) = 1.7\ J$$

Next calculate the potential energy change of the dipole in the field in terms of the current. The magnetic moment starts at an angle of 90^o with the field, and ends at an angle of 0^o:

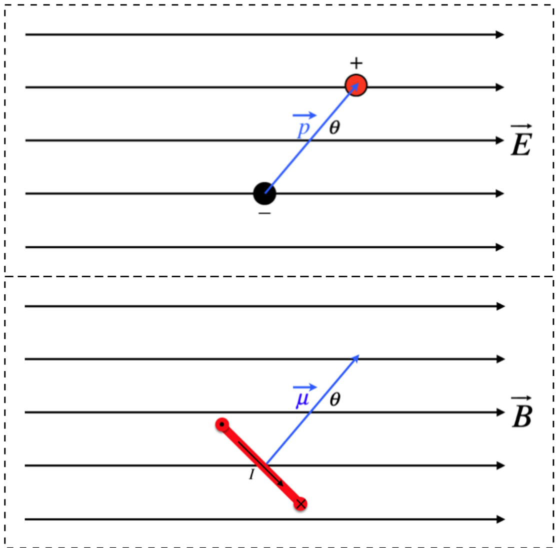
$$\Delta U_{mag} = (-\mu B \cos 0^o) - (-\mu B \cos 90^o) = -IAB = -I(0.54\ m)^2\ (1.2\ T) = -I\left(0.35\ Tm^2\right)$$

Now use energy conservation to solve for the current:

$$0 = \Delta U_{grav} + \Delta U_{mag} + \Delta KE = 1.7J - I\left(0.35\ Tm^2\right) + 0 \quad \Rightarrow \quad I = \frac{1.7\ J}{0.35\ Tm^2} = 4.9\ A$$

It is useful to make the direct comparison of these two dipoles in their respective fields with a diagram.

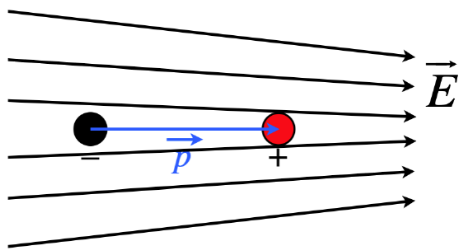
Figure 4.2.4 – Comparing Electric and Magnetic Dipoles



The two physical situations are very different, but viewing it purely in terms of the vectors, they are identical.

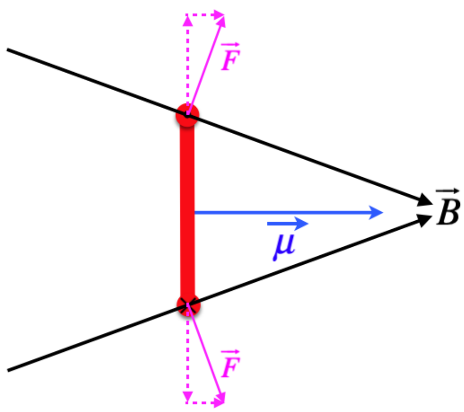
There is one last feature of dipoles we need to address – how they react to non-uniform fields. Consider an electric dipole in an electric field that gets stronger along a specific direction. When we discussed torque we used a uniform field, and found that their was no net force on the dipole. But when the field is not uniform, then half of the dipole can be in a region where the field is stronger than the region where the other half of the dipole is, resulting in more force on one part of the dipole than the other, and a net force on the dipole as a whole.

Figure 4.2.5 – Net Force on an Electric Dipole from a Non-Uniform Field



While this is clear for the electric dipole, it's not so obvious for the magnetic dipole, which doesn't have two separated charges.

Figure 4.2.6 – Net Force on an Magnetic Dipole from a Non-Uniform Field



The figure above depicts the side-view of a rectangular loop in a magnetic field that gets stronger to the right (the field lines get closer). The current is circulating clockwise when viewed from the left, so that it is going into the page in the bottom segment of the rectangle, and out of the page in the top segment. The force on the top and bottom segments must be perpendicular to the field, so there are horizontal components of this force that add together, and vertical parts that cancel, the net force being to the right, just as it is in the analogous case for the electric dipole.

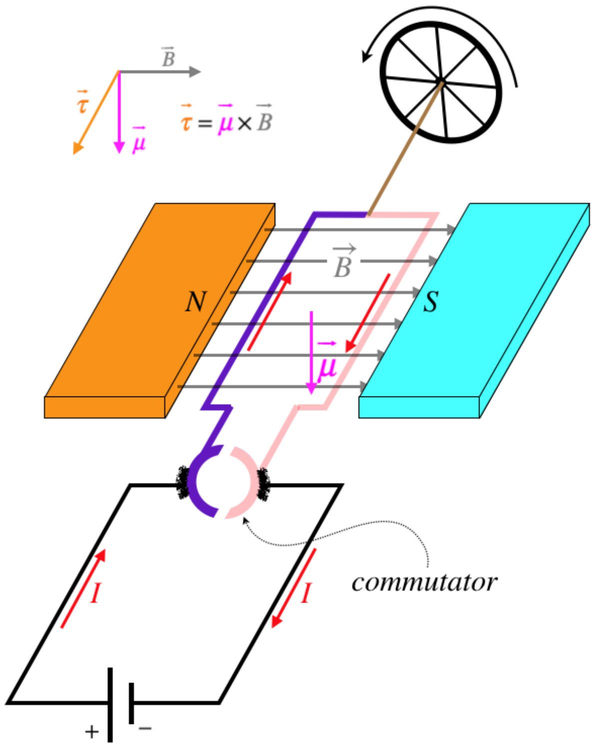
We can actually express the amount of net force exerted by the non-uniform field in terms of the rate at which the field changes along the direction of the force, using the fact that the force is the negative gradient of the potential energy. Letting the dipole moment be aligned with the field (a position into which the torque will seek to turn it), we get:

$$\begin{array}{ll}\text{electric:} & F_x=-\dfrac{\partial}{\partial x}U_E=-\dfrac{\partial}{\partial x}\left(-\overrightarrow{p}\cdot\overrightarrow{E}\right)=-\dfrac{\partial}{\partial x}(-p_xE_x)=+p_x\dfrac{\partial E_x}{\partial x}\\ \text{magnetic:} & F_x=-\dfrac{\partial}{\partial x}U_B=-\dfrac{\partial}{\partial x}\left(-\overrightarrow{\mu}\cdot\overrightarrow{B}\right)=-\dfrac{\partial}{\partial x}(-\mu_xB_x)=+\mu_x\dfrac{\partial B_x}{\partial x}\end{array} \tag{4.2.10}$$

A DC Motor

By creating a torque in a conducting loop that carries a current, we can have it do work. That is, we have the makings of an electric motor. There is one problem to overcome. As the coil turns, the direction of the magnetic moment turns with it. If the magnetic field is unchanging, then the rotation of the magnetic moment will cause the torque to reverse direction, which means the torque cause the loop to accelerate the opposite direction. This makes the coil oscillate back-and-forth, rather than rotate in a single direction. One way to fix this is to provide an emf that switches direction periodically, so that the current flips when the magnetic moment is about to provide the wrong torque. The figure below provides a simple design that fixes this problem with an unchanging emf.

Figure 4.2.7 – A Direct-Current Motor



The steady source of constant magnetic field is a bar magnet (something we will discuss in a future section). The coil rotates, but the motor includes a part called a *commutator* that has the effect of reversing the direction of the current at just the right moment. This consists of a ring with a break in it that rotates inside two rubbing connections. When the coil has rotated 90^o from the position shown in the diagram, the connection is broken briefly, and a new connection is then created which reverses the current in the wire, but keeps it going the same direction in space. In terms of the figure, when the coil turns past 90^o, the current going *away* from the wheel in the purple segment, rather than toward it, but the current is still goiing toward the wheel in the segment closest to the north magnet, so the direction of the torque remains the same. Put another way, the commutator keeps the magnetic moment of the coil pointing in a direction that always has a component downward – when it is about to start pointing upward, the current direction flips.

4.3: Magnetic Field

Field of a Magnetic Dipole

So far we have not talked about sources of magnetic fields, but even in our discussion of magnetic forces, we have not made any mention of magnetic charges that behave in magnetic fields the same way that electric charges behave in electric fields – with forces that act *along* the field lines, rather than perpendicular to them. We don't have the equivalent of Coulomb's law for two magnetic point charges, for example. Let's explore this possibility here...

Everyone has at least a passing experience with magnets – little metal disks we stick to refrigerators to hold papers we need to remember. From our experience, we know that if we put two magnets together a certain way, they stick together, and if we turn one of them around, they repel. So they clearly have a directionality to them. The closest analogy in electricity is a dipole. Indeed, if we put two dipoles end-to-end one way, they will attract, and if we turn one of them around, they will repel.

Figure 4.3.1a – Attraction of Aligned Electric Dipoles

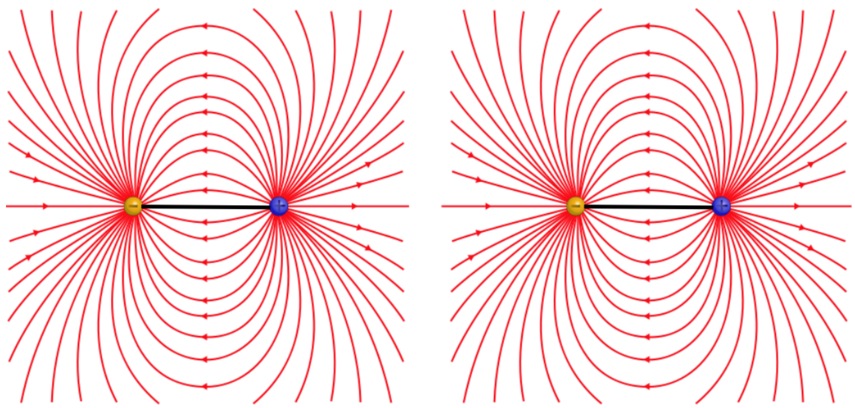
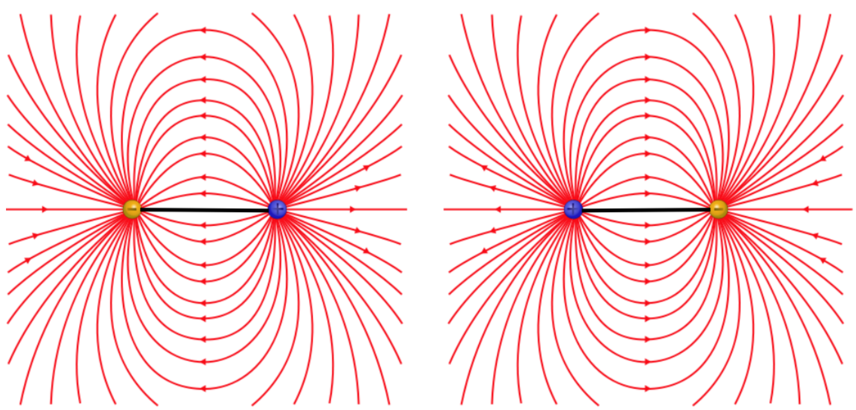


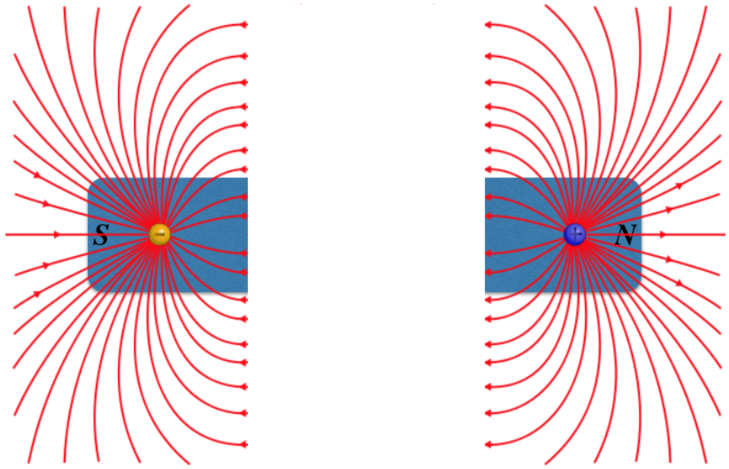
Figure 4.3.1b – Repulsion of Anti-Aligned Electric Dipoles



The attraction and repulsion occur because the there is a field created by one dipole that points in the direction outward from the positive charge, and the field gets weaker with distance, so the other dipole will feel a net force according to whichever of the two charges is closer to the dipole creating the field. In magnetism, we call the end of the magnet from which emerges the outward-going field lines the *north pole*, and the end into which the field lines converge the *south pole*. From the figures above, it's clear that the dipoles whenever like poles are brought together, and attract when opposite poles are brought together.

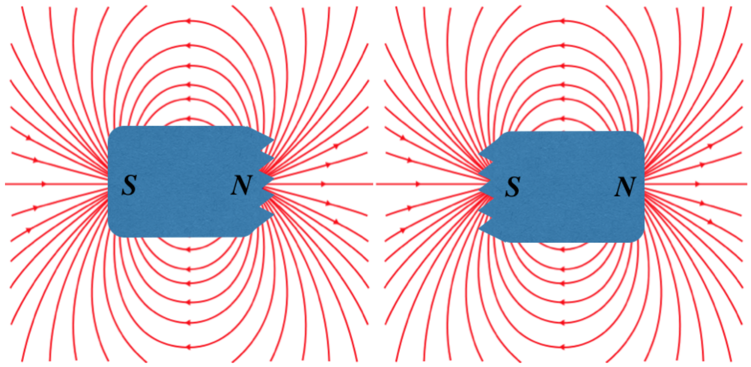
Okay, so this looks like a reasonable explanation for how magnets work, so if we want to isolate the two individual magnetic charges (a "north charge" and a "south charge"), all we have to do is cut the magnet in half, right?

Figure 4.3.2a – Isolating Magnetic Charges from a Magnet – An Attempt



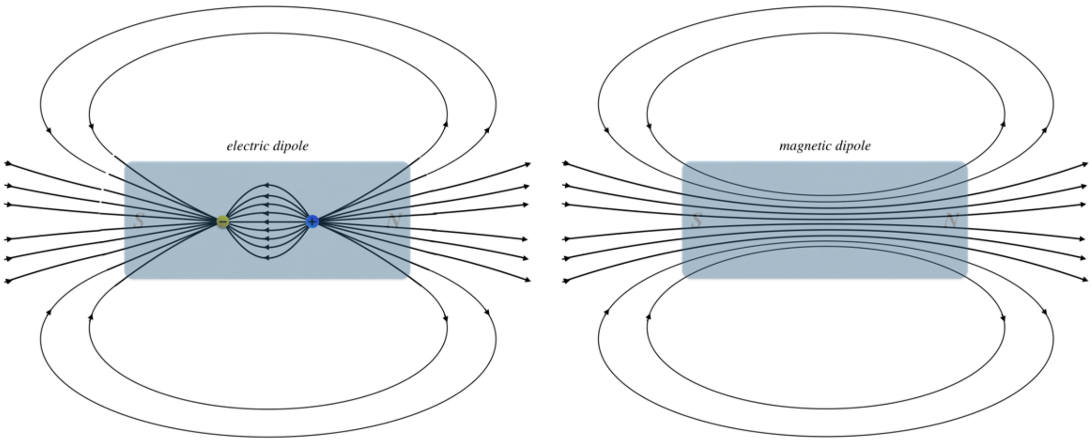
Well, try as we might, ***this never happens***. Instead, every time we cut a magnet with two poles into two pieces, we just get two more magnets with two poles!

Figure 4.3.2b – Isolating Magnetic Charges from a Magnet – What Actually Happens



If we examine the the field lines for a bar magnet closely and compare them to an electric dipole field, we see how fundamentally-different the two fields are. For the electric dipole, the field changes direction between the two poles, while for the magnetic case, the field lines continue straight through:

Figure 4.3.3 – Comparing Field Lines of Electric and Magnetic Dipoles



Outside the dipoles, the fields look the same, but they are clearly different, which we can characterize in the following way:

Magnetic field lines always form closed loops, while electric field lines begin and end on electric charges.

Put another way, unlike electric fields which form their dipole fields from two *monopoles*, there don't seem to be any magnetic monopoles. Or at least we have never been able to detect a magnetic monopole, despite many decades of experimental search for them. It turns out that electromagnetic theory doesn't exclude the possibility of the existence of these point charges of magnetism, but ultimately our theories have to agree with what we observe in experiments, so at least for the moment (and for the duration of this class), we will maintain the position that they simply don't exist.

Gauss's Law for Magnetism

The revelation that our theory of magnetism doesn't include individual magnetic charges has an immediate consequence for the magnetic equivalent of Gauss's law. With magnetic field lines always forming closed loops, any field line that penetrates a Gaussian surface going in one direction (say going into the volume bounded by the surface) must later emerge from that closed surface later in order to form the closed loop. If there is a field line exiting a surface for every field line that enters it, then the net flux must necessarily always be zero. Of course, from Gauss's law, this means that there can never be any charge enclosed, and this makes sense, given that there is no magnetic charge!

Mathematically, we express this Gauss's law for magnetism in either integral or local form:

$$\oint \overrightarrow{B} \cdot d\overrightarrow{A} = 0 \;, \qquad \overrightarrow{\nabla} \cdot \overrightarrow{B} = 0 \tag{4.3.1}$$

Field of a Moving Point Charge

When we first started discussing magnetism, we noted a force between two current-carrying wires. From there, we focused on the fact that a magnetic field affects only *moving* electric charges, but it should be equally clear that the *source* of a magnetic field must also be moving electric charges. One might object that we just said that magnetic fields don't have point sources, so what difference does it make that we insist that the point source be moving? We will see that this makes all the difference, because this leads to a field that doesn't point directly toward or away from that charge – the direction of the field is determined by the direction of the velocity vector.

As different as the magnetic field is from the electric field, there are still so many striking similarities that it is useful to describe the features of the magnetic field from a moving point charge in parallel with the Coulomb electric field. This magnetic analog of the Coulomb field is called the *law of Biot & Savart*.

- In the Coulomb case, we started with the fact that the field strength is proportional to the magnitude of the charge emitting the field. In the magnetic case, the field strength is also proportional to the magnitude of the charge, but since the charge must also be moving, it turns out that the field strength is also proportional to the charge's *speed*. This agrees with the observation that there is no magnetic field if the charge is stationary.
- Next we consider how the strength of the field weakens with distance from the source. In this case, the two fields behave identically – with an inverse-square law.
- The direction is where these two fields differ the most. In the Coulomb case, the field points directly toward or away from the point charge. Put another way, if the source charge is at the origin, then the electric field at a position in space described by a position vector $\overrightarrow{r}$ points in a direction that is parallel to that position vector. The magnetic field, by contrast points *perpendicular* to that position vector. This doesn't narrow down the direction, however, as there is an entire plane that is perpendicular to this vector. This is where the velocity vector direction comes in – the magnetic field is also perpendicular to the direction defined by the velocity vector. We already know a way of expressing a vector that is perpendicular to two other vectors at the same time – it must be parallel to the cross product of those two vectors.

Putting all these features together, and including physical constants to make the units work out correctly, we get the following summary:

$$\begin{array}{ccc} \text{feature} & \text{Coulomb} & \text{Biot-Savart} \\ \text{field source:} & \left|\overrightarrow{E}\right| \propto q \;\; \text{(charge)} & \left|\overrightarrow{B}\right| \propto q\left|\overrightarrow{v}\right| \;\; \text{(moving charge)} \\ \text{field strength with distance:} & \left|\overrightarrow{E}\right| \propto \dfrac{1}{r^2} & \left|\overrightarrow{B}\right| \propto \dfrac{1}{r^2} \\ \text{direction:} & \overrightarrow{E} \parallel \overrightarrow{r} & \overrightarrow{B} \parallel \overrightarrow{v} \times \overrightarrow{r} \\ \text{physical constant:} & \overrightarrow{E} = \left(\dfrac{1}{4\pi\epsilon_o}\right)\dfrac{q\overrightarrow{r}}{r^3} & \overrightarrow{B} = \left(\dfrac{\mu_o}{4\pi}\right)\dfrac{q\overrightarrow{v}\times\overrightarrow{r}}{r^3} \end{array} \tag{4.3.2}$$

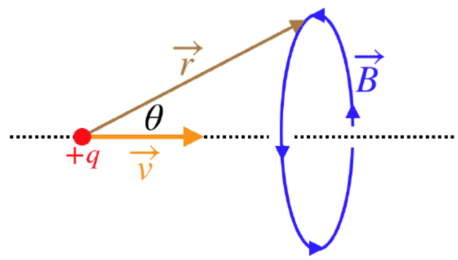
The physical constant that makes the units work out for the force is called the *permeability of free space*:

$$\mu_o = 4\pi \times 10^{-7} \, \frac{T \cdot m}{A}$$

Yes, that is exactly 4π in that constant. The unit of Tesla was constructed to come out to Newtons, which explains why the 4π cancels-out in Biot-Savart's law. One might wonder why we bother to introduce the constant this way at all, and the answer to this question will become clear later. Right now the short answer is that it will parallel very closely the role that ϵ_o plays in electricity.

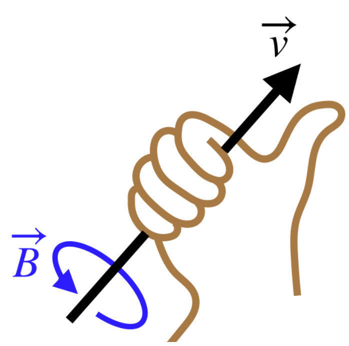
While it is not obvious from the final form of the equation for the magnetic field, the resulting field is a circle centered at the line passing through the charge along the direction of motion:

Figure 4.3.4 – Magnetic Field of a Moving Point Charge



Rather than using the right-hand-rule for the cross-product $\overrightarrow{v} \times \overrightarrow{r}$ (which gives the direction of the magnetic field at a specific point in space), we can get a bigger-picture idea of the magnetic field lines by using a different right-hand-rule: Point the thumb of the right hand in the direction of motion of the charge, and the magnetic field direction everywhere in space forms closed circles around the line of motion in the direction that the fingers curl.

Figure 4.3.5 – Right-Hand-Rule for Magnetic Field



Field of a Current-Carrying Wire

It is far more common to have physical situations where a magnetic field is created by a current-carrying wire than by a point charge. Fortunately, we already know how to convert from moving point charges to current elements:

$$I \, \overrightarrow{dl} \leftrightarrow dq \, \overrightarrow{v} \tag{4.3.3}$$

We therefore get this for a line of current from the law of Biot & Savart:

$$\vec{B} = \int d\vec{B} = \int \left[\left(\frac{\mu_o}{4\pi} \right) \frac{I}{r^2} \vec{dl} \times \hat{r} \right] = \frac{\mu_o}{4\pi} \int \frac{I \vec{dl} \times \vec{r}}{r^3} \tag{4.3.4}$$

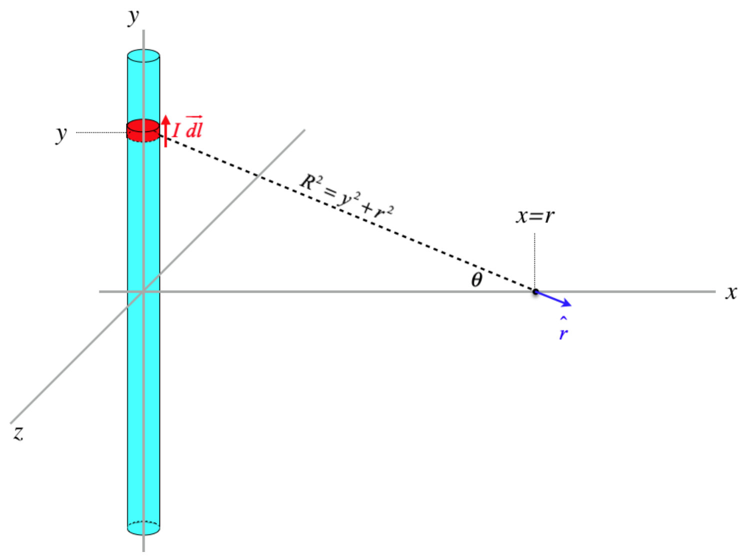
We can now use this result to continue building our "toolbox" of reusable solutions of common physical sources.

4.4: Sources of Magnetic Fields

Magnetic Field of a Long Straight Wire

We begin by computing the field of a long-straight wire that carries a current I. Aside from the vectors, the procedure follows almost exactly the same path as the case of the electric field of a long line of charge.

Figure 4.4.1 – Calculating Magnetic Field of Long, Straight Wire



One of the key differences between computing magnetic fields and electric fields is that while we were able to use symmetry to help us solve for components of the electric field, in the case of the magnetic field, this is much harder to do, and is much safer to just get all the vectors right and trust vector math thereafter. We could have used this "trust the vector math" approach for the electric field as well, of course, but the necessity of using it in cases where cross-products are involved becomes quickly apparent.

Okay, we start by expressing all the relevant quantities in terms of our chosen coordinate system:

$$\overrightarrow{dl} = dy\;\hat{j} \qquad \hat{r} = \cos\theta\;\hat{i} - \sin\theta\;\hat{j} \qquad \cos\theta = \frac{r}{\sqrt{y^2+r^2}} \tag{4.4.1}$$

Next, write down Biot-Savart's law for the current element, and simplify:

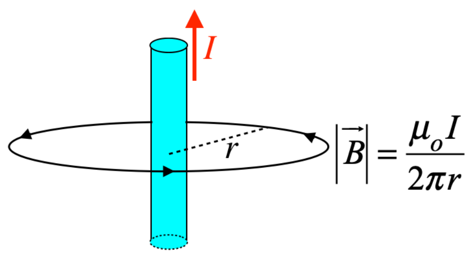
$$\begin{aligned} d\overrightarrow{B} &= \left(\frac{\mu_o}{4\pi}\right)\frac{I}{R^2}\overrightarrow{dl}\times\hat{r} \\ &= \left(\frac{\mu_o I}{4\pi}\right)\frac{dy}{y^2+r^2}\hat{j}\times\left(\cos\theta\;\hat{i}-\sin\theta\;\hat{j}\right) \\ &= \left(\frac{\mu_o I}{4\pi}\right)\frac{dy}{y^2+r^2}\left(\cos\theta\;\;\cancelto{-\hat{k}}{\hat{j}\times\hat{i}} - \sin\theta\;\;\cancelto{-0}{\hat{j}\times\hat{j}}\right) \\ &= \left(\frac{\mu_o I}{4\pi}\right)\frac{dy}{y^2+r^2}\left(\frac{r}{\sqrt{y^2+r^2}}\right)\left(-\hat{k}\right) \\ &= \left(\frac{\mu_o I\;r}{4\pi}\right)\frac{dy}{\left(y^2+r^2\right)^{\frac{3}{2}}}\left(-\hat{k}\right) \end{aligned} \tag{4.4.2}$$

All that remains is to add up the contributions to the field from all the current elements, which means integrating this from $y=-\infty$ to $y=+\infty$:

$$\begin{aligned}\vec{B} &= \left(-\hat{k}\right)\left(\frac{\mu_o I\ r}{4\pi}\right)\int\limits_{-\infty}^{+\infty}\frac{dy}{\left(y^2+r^2\right)^{\frac{3}{2}}} \\ &= \left(-\hat{k}\right)\left(\frac{\mu_o I\ r}{4\pi}\right)\left[\frac{1}{r^2}\ \frac{y}{\sqrt{y^2+r^2}}\right]_{-\infty}^{+\infty} \\ &= \left(-\hat{k}\right)\left(\frac{\mu_o I}{4\pi r}\right)[2] \\ &= \left(\frac{\mu_o I}{2\pi r}\right)\left(-\hat{k}\right)\end{aligned} \tag{4.4.3}$$

The resemblance the magnitude of this field bears to that of the electric field (Equation 1.5.2) is interesting, though not all that surprising, given that both fields weaken with distance from the source according to an inverse-square law. The direction of the magnetic field vector is tangent to a circle centered at the line of the current, and circles around the current line.

Figure 4.4.2 – Magnetic Field Circulates Around the Long, Straight Wire

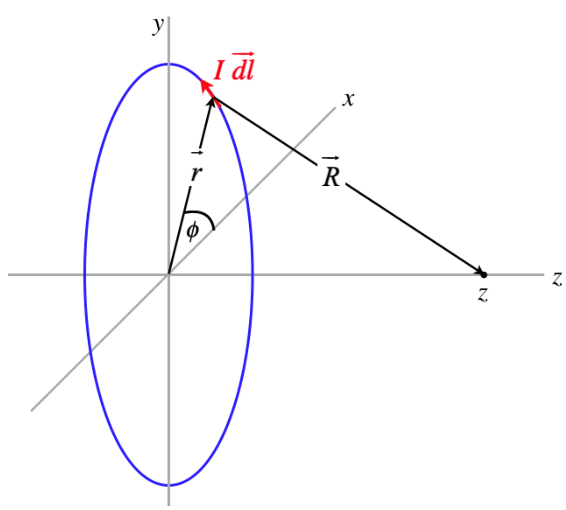


As with the electric field, the magnetic field obeys superposition, which means we can combine the result of this physical situation with others to get a net magnetic field. It is also worth noting that both the moving point charge and the long, straight wire yield magnetic fields whose line close back on themselves (form closed loops) – in nether case does a field emanate out of or into the source. There are no magnetic monopole fields.

Field of a Loop

Another useful field to know is that which points along the axis of a circular loop of current. The method is essentially the same as above, but the coordinate system used is different, which leads to a little bit more complicated vector manipulation.

Figure 4.4.3 – Calculating Magnetic Field on the Axis of a Circular Loop of Current



Again start by expressing quantities in terms of the coordinates we have set up. We can again write everything in terms of the ijk unit vectors, but this time we can do it a bit differently. First we have the magnitude of the segment of wire:

$$\left|\vec{dl}\right| = r\ d\phi \tag{4.4.4}$$

Next we note that tail-to-head vector addition gives:

$$\overrightarrow{r}+\overrightarrow{R}=z\hat{k} \quad\Rightarrow\quad \overrightarrow{R}=-\overrightarrow{r}+z\hat{k} \tag{4.4.5}$$

Biot-Savart's law gives:

$$d\overrightarrow{B}=\frac{\mu_o}{4\pi}\frac{I\overrightarrow{dl}\times\overrightarrow{R}}{R^3} \tag{4.4.6}$$

Before we can integrate, we have to resolve the vector products. Looking at the diagram, we can see that the current element $\overrightarrow{dl}$, the position vector of the current element $\overrightarrow{r}$, and the unit vector $\hat{k}$ are all mutually orthogonal, making $\overrightarrow{dl}\times\overrightarrow{r}$ parallel to $\hat{k}$, and $\overrightarrow{dl}\times\hat{k}$ parallel to $\overrightarrow{r}$. This allows us to use the right-hand rule to complete these products:

$$\overrightarrow{dl}\times\overrightarrow{R}=\overrightarrow{dl}\times\left(-\overrightarrow{r}+z\hat{k}\right)=r\;dl\;\hat{k}+z\;dl\;\hat{r} \tag{4.4.7}$$

Putting this result into the integral and noting that magnitudes of the vectors $\overrightarrow{r}$ and $z\,\hat{k}$ are constant in the integral, and satisfy $R^2=r^2+z^2$, we get:

$$\overrightarrow{B}=\frac{\mu_o I}{4\pi}\int\frac{\overrightarrow{dl}\times\overrightarrow{R}}{R^3}=\frac{\mu_o I}{4\pi\left(r^2+z^2\right)^{\frac{3}{2}}}\int\left[r\;dl\;\hat{k}+z\;dl\;\hat{r}\right] \tag{4.4.8}$$

While the magnitude of $\overrightarrow{r}$ doesn't change over the integral, its direction does change, so we have to write the unit vector $\hat{r}$ in terms of the coordinates to do the integral of the second term. Let's do each integral separately. The first is straightforward, since the integral of just dl is simply the circumference of the circle

$$\begin{gathered}\frac{\mu_o I r}{4\pi\left(r^2+z^2\right)^{\frac{3}{2}}}\hat{k}\int dl=\frac{\mu_o I r^2}{2\left(r^2+z^2\right)^{\frac{3}{2}}}\hat{k}\\ \frac{\mu_o I z}{4\pi\left(r^2+z^2\right)^{\frac{3}{2}}}\int dl\hat{r}=\frac{\mu_o I z}{4\pi\left(r^2+z^2\right)^{\frac{3}{2}}}\int_0^{2\pi}r\;d\phi\left(\cos\phi\;\hat{i}+\sin\phi\;\hat{j}\right)=0\end{gathered} \tag{4.4.9}$$

The second integral just ends up vanishing, giving the result for a magnetic field along the axis of a loop of radius r a distance z from the plane of the loop:

$$B=\frac{\mu_o I r^2}{2\left(r^2+z^2\right)^{\frac{3}{2}}} \tag{4.4.10}$$

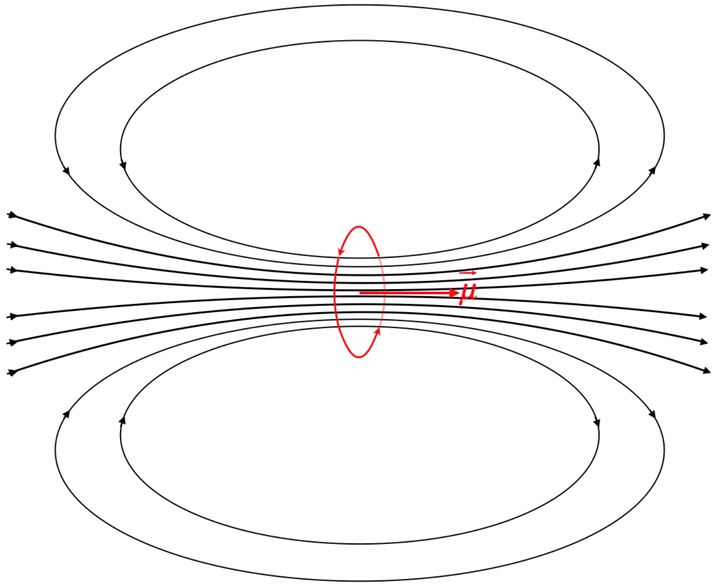
If we are only interested in the field at the *center* of the loop, we plug in $z=0$ to get the simple result:

$$B=\frac{\mu_o I}{2r} \tag{4.4.11}$$

The direction of $\hat{k}$ shows us yet another shortcut for using the right-hand-rule for the field along the axis (*and only along the axis!)* of a loop: Curl the fingers of the right hand such that they trace the circulation of the current around the loop, and the thumb points the direction of the field.

We have already talked about a loop as a magnetic dipole which interacts with fields that are present, and here (as in the case of the electric dipole), we see that the dipole also emits a field, and this field – like the electric dipole field – gets weaker as the inverse cube of the distance (which in this case is measured by z). Also, like the electric dipole, the field along its axis points in the direction of the dipole moment:

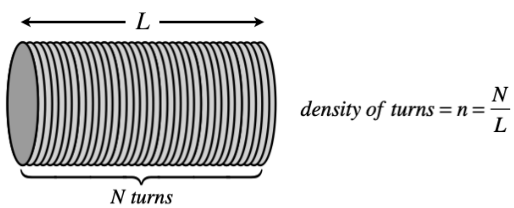
Figure 4.4.4 – Magnetic Dipole Field of a Loop



Field of a Solenoid

It is possible to stack lots of individual dipoles on top of each other to create a long tube called a *solenoid*. Such a device consists of a number of turns in the coil N, and a length L, resulting in what will be the critical measure, the *turn density:*

Figure 4.4.5 – A Solenoid



How do we compute the field for such an object? Well, first of all, we need to specify what field we want. Like the loop, we will only look on the axis. But we will also simplify it further by assuming we are looking at a point on the axis *inside* the solenoid far from the ends (so essentially it has an infinite length, though the turn density is of course finite).

We treat this as a collection of an infinite number of loops. If we pick an origin (which we can place anywhere along the infinite axis), then we have the field at that point by a loop at a position z on the axis is given by Equation 4.4.10. Then we need to add up the field contributions at the origin due to all of the loops. The problem is, there is not a loop at every point along the z-axis. With a turn density of n, the number of turns in a tiny slice dz would be $n\;dz$ The total current in that slice would then be this number multiplied by the current through the wound wire (which we will call I):

$$\text{current in a } dz \text{ slice located at } z = I\;n\;dz \tag{4.4.12}$$

Plugging this into Equation 4.4.10 for the current, gives the tiny contribution to the field by the slice, and adding them all up gives the field. We are not given the radius of the solenoid, but we will call it r, and we'll see that it isn't relevant!

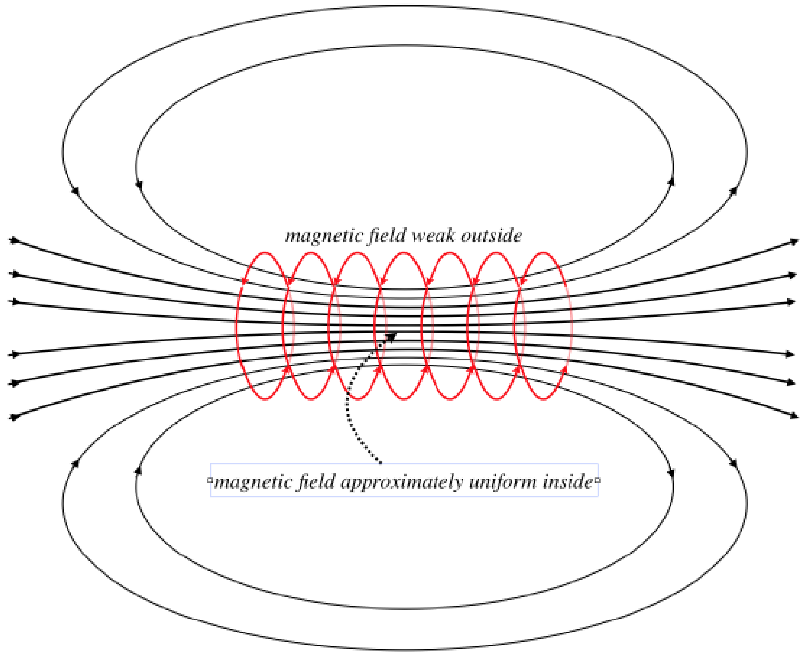
$$B=\int\limits_{-\infty}^{+\infty}\frac{\mu_o\left(n\;I\;dz\right)r^2}{2\left(r^2+z^2\right)^{\frac{3}{2}}}=\frac{\mu_o\;n\;I\;r^2}{2}\int\limits_{-\infty}^{+\infty}\frac{dz}{\left(r^2+z^2\right)^{\frac{3}{2}}}=\frac{\mu_o\;n\;I\;r^2}{2}\left[\frac{z}{r^2\sqrt{r^2+z^2}}\right]_{-\infty}^{+\infty}=\mu_o\;n\;I \tag{4.4.13}$$

There are a few particularly interesting aspects of the fields of solenoids, which are not immediately evident from this solution, but which we will state without proof (for now – we have another tool to use later that makes this easier):

- The field within the solenoid doesn't change much (it is pretty much *uniform*). This basically comes from the fact – as we found here – that the field on the axis doesn't depend upon the radius of the solenoid.
- The field just outside the solenoid (on its side, not the end) is very weak (basically it is *zero*).

- The field looks just like that of a bar magnet, but it can be turned on and off by switching the current on or off.

Figure 4.4.6 – Solenoid Magnetic Field



Digression: Electromagnets

Solenoids have lots of practical uses, a common one being something known as an "electromagnet." For example, junk yards use these to move large chunks of scrap metal. Obviously the ability to cut the current to turn off the magnetic field is key here. If the crane used a permanent magnet, it wouldn't be able to let go of the crushed car. Another application is for fire doors. Imagine large doors held open in hallways of a building with electromagnets, and if a fire breaks out, the power is cut and the doors close, hopefully slowing the spread of the fire. Gates where you are "buzzed-in" are held shut by a latch that is released with the activation of an electromagnet that draws the latch back with a magnetic field.

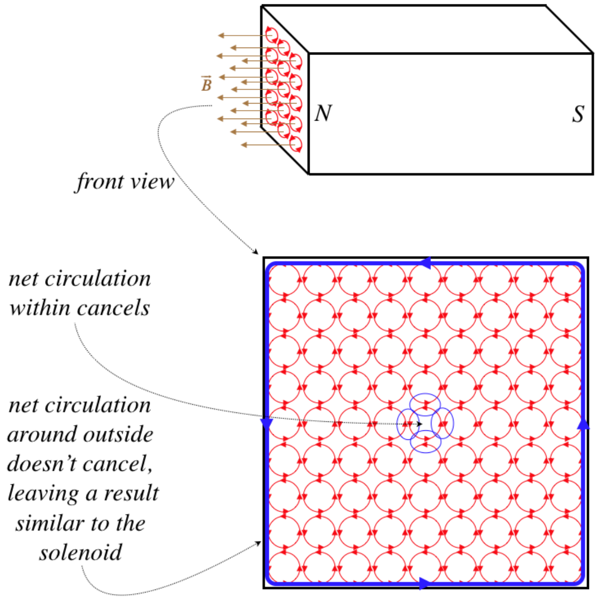
Magnetic Materials

At last we can discuss elements of magnetism that we have been aware of since we were young kids – the properties and behavior of bar magnets. As with any phenomenon that requires an understanding of what is going on at a microscopic level, magnetism inside of materials like bar magnets is very complicated. We'll look at a greatly-simplified version of it here, but keep in mind that a fuller understanding can only be achieved through quantum theory.

We now know that there are no "magnetic particles" comprising a bar magnet – its magnetic field can only be created by moving charges. But unlike an electromagnet, bar magnets are not plugged into some emf source, so where does the moving charge come from? The atoms that comprise the material of course include lots of charges, and these charges are moving in manners that resemble magnetic dipoles – electrons are orbiting nuclei in more-or-less circular loops, and electrons also have a quantum-mechanical property called "spin" that gives them their own magnetic moments as well.

We will not concern ourselves too much with the specific source of the magnetic moments of these particles, but we will instead just focus on the fact that each particle has its own magnetic moment. In the case of a bar magnet, these dipoles tend to be permanently aligned. This property is called *ferromagnetism*. Only a few materials have this property: iron (thus "ferro"), nickel, cobalt, many of their alloys, and some of the rare earth metals. [*Technically, these alignments occur in chunks, called "domains," within the magnet, and the degree to which the magnet is magnetized is determined by how much certain domains "swallow-up" others, creating more broadly-coordinated alignments of dipoles.*]

Figure 4.4.7 – Dipoles in a Bar Magnet



It should be clear how two bar magnets attract each other. One is a magnetic field source, and the other is a magnetic dipole that experiences the non-uniform field of the other. The field diverges as it emerges from one magnet, and the dipole of the other magnet, if the poles are aligned, reacts by feeling a force in the direction where the field gets stronger, following the mechanism depicted in Figure 4.2.6.

But this doesn't explain why a magnet sticks to a refrigerator when the refrigerator metal is not itself magnetized. The answer is that some metals, while their particle dipole moments are not permanently aligned, have the property that their particles are free to rotate their magnetic moments. When an external field is applied, their particles then align, making them magnetized. When the field is removed, the random alignments return. This property is known as *paramagnetism*.

A few general comments to close this subject:

- Ferromagnetism relies primarily upon the spin source of magnetic moment, and very little on the orbital source, while paramagnetism relies upon both.
- Ferromagnetic materials remain magnetized after a strong applied magnetic field aligns the domains, which remain aligned thanks to anomalies in the crystal structure which "snag" the domains and hold them in an aligned orientation.
- Ferromagnets can be demagnetized ("degaussed") by relieving these snags. This is most easily done by raising the temperature to a critical temperature known as the *Curie temperature*, at which magnetic domain "snags" are no longer possible and the substance has zero ferromagnetism. Other methods for degaussing include applying rapidly-changing magnetic fields (which "shake" the domains into random orientations) and pounding on the magnet so that vibrations cause the domains to un-snag.

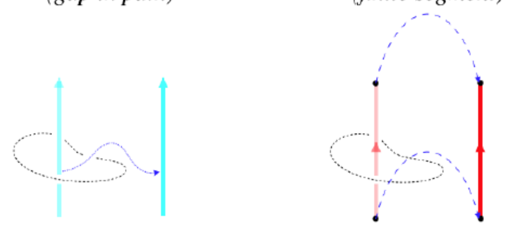
4.5: Ampére's Law

A Magnetic Analog to Gauss's Law for Electricity

We have already stated that Gauss's law for magnetic field is trivial, as there are no magnetic monopoles, but it turns out that there is a separate mathematical law that works for magnetic sources in the same way that Gauss's law works for electric sources. It incorporates "enclosed" sources, and allows us to use symmetrical situations to solve for fields using methods simpler than integrating Biot-Savart's law.

This magnetic version of the electrical Gauss's law is called *Ampére's law*, and since it can't involve enclosed point sources, it instead deals with lines of current, which either circle back on themselves to form a closed circuit, or are infinitely-long (and circle-back on themselves at infinity). But how do we "enclose" a line of current? In the case of charge, it was enclosed if there was no way to remove the charge from the Gaussian surface without breaking through the surface (i.e. the surface has no holes in it). In the case of Ampere's law, we consider a current to be enclosed by an imaginary closed path – called an *Ampérian circuit* – rather than a surface. Such a current is enclosed when there is no way to move the line of current out without it breaking through the Ampérian circuit.

Figure 4.5.1 – Defining an Enclosed Current



There are two ways that a current won't be enclosed: If there is a break in the path around the wire, so that the wire can slide through it, or is the segment of wire is finite in length and does not form a closed loop. If the loop of wire is closed, then the Ampérian circuit is "linked" with it, and the current is enclosed. If the wire is infinitely long, the wire similarly cannot escape the Ampérian circuit without breaking through it, so its current is also enclosed. [*Mathematically, we generally define a current to be "enclosed" by a closed path if it pierces* ***every possible*** *surface that is bounded by the closed path. Clearly there are stretched surfaces we can construct with the closed path as it border that do not allow a finite-length segment to pierce it.*]

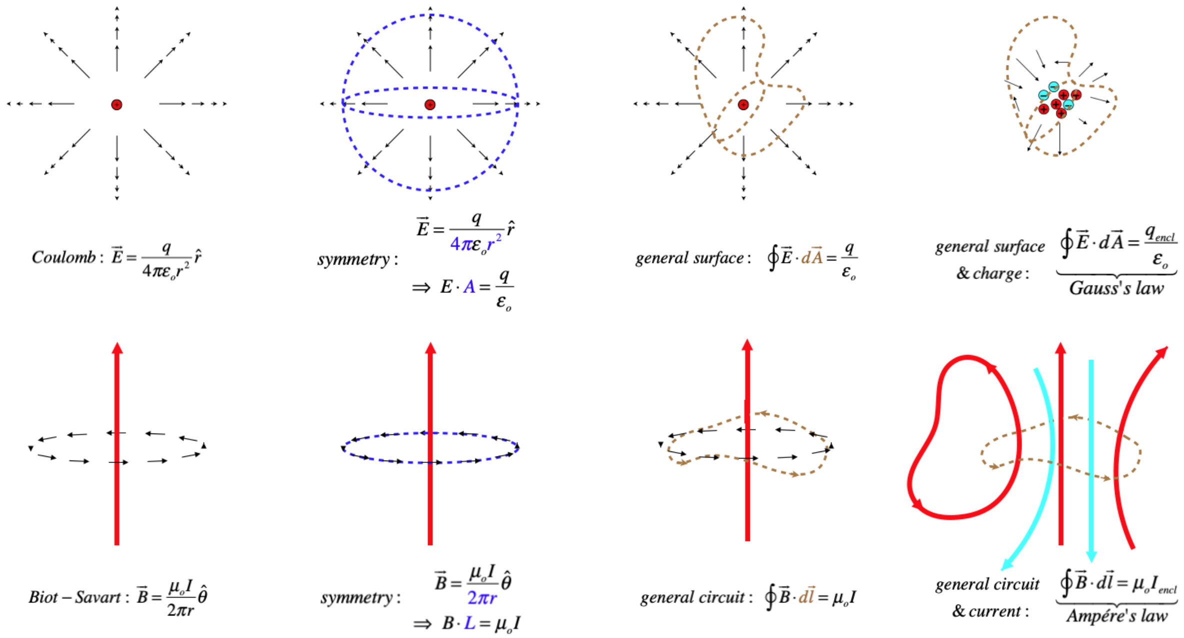
With this definition of enclosed current in place, we have Ampére's law:

$$\oint \overrightarrow{B} \cdot \overrightarrow{dl} = \mu_o I_{enclosed} \tag{4.5.1}$$

The integral is performed along *any* Ampérian circuit that goes completely around the enclosed current, in the same way that the integral for Gauss's law works for any closed surface that encloses the charge. All of the properties of Gauss's law have analogous properties for Ampére's law:

- Enclosed is defined in terms of ability to remove charge/current from surface/closed path without breaking through the enclosure.
- The sign of the charge inside a Gaussian surface is related to the positive direction of the area vector. If the total flux is in the negative direction (opposite to the area orientation), then the enclosed charge is negative. Similarly, we define a positive *direction of circulation* for an Ampérian circuit, and if the direction of the magnetic field of the current has the same circulation orientation as that of the Ampérian circuit, then that current is "positive," otherwise it is "negative."
- There can be both positive and negative charges/currents enclosed in a surface/closed path, and these are combined to give a net charge/current.
- We can use symmetry to solve for a field. This usually means that the field is either parallel or perpendicular to the surface/closed path, and that it has a constant magnitude on that surface/closed path.
- The shape of the surface/closed path is not relevant, as long as it is closed.

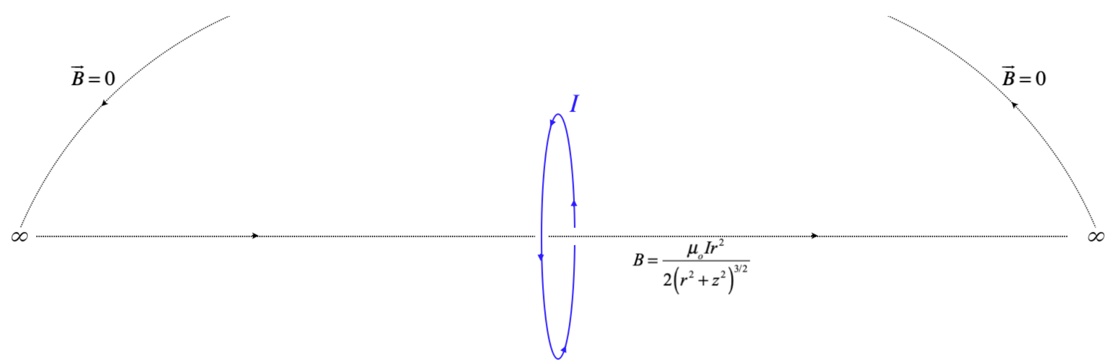
Figure 4.5.2 – Comparing Gauss's Law to Ampére's Law



Applications of Ampére's Law

The applications of Ampére's law are much like those of Gauss's law – symmetry is used to compute fields. Before we do any of these, let's confirm a result we already have – the field on the axis of a circular loop of radius r (Equation 4.4.10). We need to choose a useful circuit – it needs to enclose the current, and it needs to have a direction that matches well with the direction of the magnetic field that we know from symmetry. To this end, we will choose an Ampérian circuit that follows the z-axis from $-\infty$ to $+\infty$, and from there circulates back around on itself at infinity. The integral of this infinitely-distant segment of the closed path will vanish, because the field weakens to zero this far away, so all that is left is the integral along the z-axis.

Figure 4.5.3 – Ampérian Circuit for Loop



The magnetic field on the axis is in the same direction as the circuit along the z axis, and while we don't know the angle between the field and the direction of the path elsewhere, it doesn't matter, because the field vanishes. So the closed-path line integral gives:

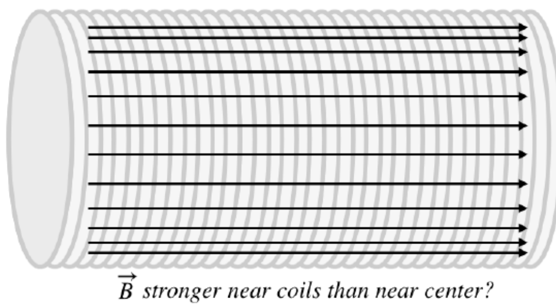
$$\oint \vec{B} \cdot \vec{dl} = \int\limits_{z-axis} Bdz + \int\limits_{R=\infty} \vec{B} \cdot \vec{dl} = \frac{\mu_o I r^2}{2} \int\limits_{-\infty}^{+\infty} \frac{dz}{\left(r^2+z^2\right)^{\frac{3}{2}}} + 0 = \frac{\mu_o I r^2}{2} \left[\frac{z}{r^2\sqrt{r^2+z^2}} \right]_{-\infty}^{+\infty} = \mu_o I \qquad (4.5.2)$$

Since the enclosed current within this Ampérian circuit is in fact the current in the circular loop, and since this current has a positive orientation for this closed path (it is coming out of the page and the path is counterclockwise, satisfying the right-hand-rule), Ampére's law is confirmed for this case.

Next let's use Ampére's law in a more traditional fashion – to compute the magnetic field in a symmetric physical situation. We'll look at the field of a solenoid. The solenoid has a high degree of symmetry, mainly thanks to our assumption that it is essentially infinite in length. Consider the direction of the field inside the solenoid. There is no reason to expect that it would diverge from being parallel with the axis anywhere, because then what would make the point where this divergence occurs so special? It looks like every other point in the infinite solenoid, so the field direction should be the same as every other point in the solenoid. Furthermore, it should not get any stronger along the axis – why would the field be a greater magnitude at one point along the axis than any other?

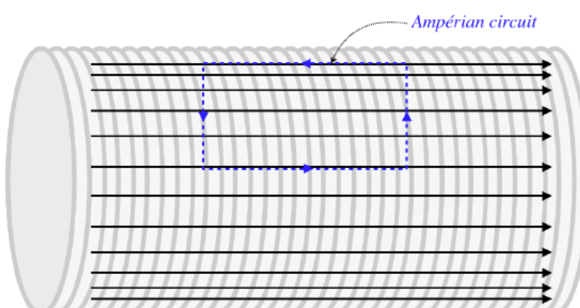
Of course, knowing that the field inside the solenoid is parallel to the axis everywhere and equal strength as a function of its position along the axis inside the solenoid doesn't tell us that the field strength is also uniform as a function of distance from the axis – perhaps it gets stronger as the position gets closer to the wire coils? For example, the field may look like this:

Figure 4.5.4 – Speculation About the Magnetic Field Inside a Solenoid



We can check this speculation using Ampére's law. We do this by constructing a rectangular Ampérian circuit inside the solenoid that has two sides parallel to the field, and two sides perpendicular to the field.

Figure 4.5.5 – An Ampérian Circuit for the Interior of the Solenoid



Now we apply Ampére's law. First of all, there is no electric current flowing through this closed path, so the integral around the path will be zero:

$$\oint \overrightarrow{B}\cdot\overrightarrow{dl}=\mu_o I_{enclosed}=0 \tag{4.5.3}$$

Second, the magnetic field is perpendicular to two of the sides of the rectangle, so the dot product of the field with the displacements during that portion of the circuit is zero, giving zero integral along the vertical segments. And the field strength doesn't change along the direction of the axis, while the field is parallel to these parts of the path, so the dot products and integrals are easy to perform:

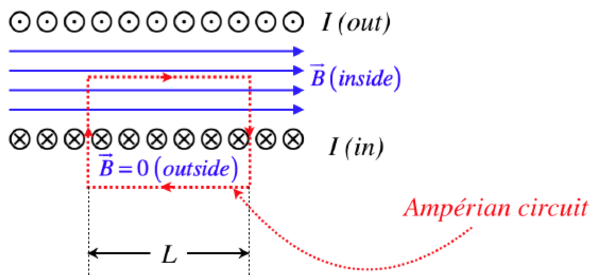
$$\begin{aligned} \int_{vertical} \overrightarrow{B}\cdot\overrightarrow{dl} &= 0 \\ \int_{top} \overrightarrow{B}_{top}\cdot\overrightarrow{dl} &= \int\left(-B_{top}\,dl\right) &= -B_{top}L \\ \int_{bottom} \overrightarrow{B}_{bottom}\cdot\overrightarrow{dl} &= \int\left(+B_{bottom}\,dl\right) &= +B_{bottom}L \end{aligned} \tag{4.5.4}$$

Adding all four contributions to the integral and setting it equal to zero (no enclosed current) gives us the result that $B_{top} = B_{bottom}$, which means that the field strength at different distances from the axis inside the solenoid is actually the same. The magnetic field in the interior of a solenoid is completely uniform.

Notice that a rectangular Ampérian circuit parallel to the axis drawn *outside* the solenoid must give us the same result: The magnetic field must be uniform out there as well. But that means that the field is the same strength infinitely-far from the solenoid as it is just outside of it. Naturally the field is zero infinitely-far away, which means that we can conclude that the magnetic field is zero everywhere outside the solenoid.

Finally, we can also use Ampére's law to determine the strength of the uniform magnetic field inside the solenoid, without going through the integration of an infinite number of loops, as we did in the previous section. We accomplish this by constructing a rectangular Ampérian circuit parallel to the axis with one side inside the solenoid, and one outside of it.

Figure 4.5.6 – An Ampérian Circuit to Compute Field Inside a Solenoid



As before, we see that the contributions to the integral by the vertical sides of the rectangular path are zero, because inside the solenoid the field is perpendicular to the path (and outside the solenoid the field is zero). This time, however, the contribution by one of the horizontal sides (the one outside the solenoid) is also zero, since there is no field. The side of the rectangle inside the solenoid is easy to integrate, so putting them all together into Ampére's law gives:

$$\mu_o I_{enclosed} = \oint \overrightarrow{B} \cdot \overrightarrow{dl} = BL\,(inside) + 0\,(vertical) + 0\,(outside) \tag{4.5.5}$$

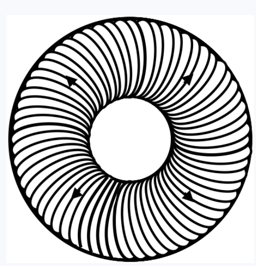
Now we have to determine the enclosed current. From the diagram, we can see using the right-hand-rule for the Ampérian circuit that the current going into the page is positive. The diagram shows 6 wires enclosed, but we will call this a more generic N. If the current flowing through the wire is I, then the enclosed current is then $+NI$. Plugging this in above gives:

$$\mu_o NI = BL \quad \Rightarrow \quad B = \mu_o \frac{N}{L} I = \mu_o nI \tag{4.5.6}$$

This matches the result we found previously (Equation 4.4.13). And as was the case with Gauss's law for electric fields, the mathematical gymnastics required is greatly reduced using this method.

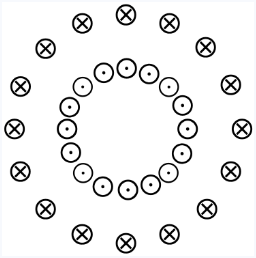
Example 4.5.1

Consider a toroidal solenoid. This is a solenoid that is finite in length, and has had its axis bent such that the two open ends are closed upon each other, forming a classic doughnut (or bagel) shape. When we look at this device, from our perspective, the current is passing through the wires coming out of the page from the center of the toroid, and going into the page on the outside of the toroid (see the diagram below). use Ampére's law to find the field inside and outside the solenoid.



Solution

It's best to start by looking at a cross-section of the solenoid, slicing through the "bagel" parallel to the plane of the page . This shows wires on the inside with current coming out of the page, and wires on the outside with current going into the page.

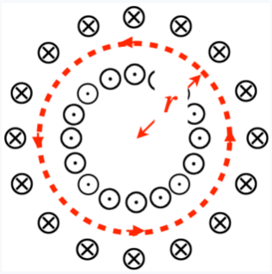


From symmetry, we can conclude that the magnetic field at any fixed distance from the center is the same magnitude, though the field strength may vary when that distance is changed. Also, symmetry demands that the field direction will have to be tangent to a circle in the plane of the page and concentric to the center of the toroid.

Constructing an Ampérian circuit that is a circle in the plane of the page will then be parallel to whatever field is present anywhere, making the integral easy to do. If the radius of the circular path is r*, then:*

$$\oint \vec{B} \cdot \vec{dl} = \oint B\, dl = B \oint dl = B\,(2\pi r)$$

If we choose the circuit to be outside the toroid (or inside the doughnut hole), then the enclosed current is zero, which tells us that the magnetic field outside the toroid is zero. Like a straight solenoid, a toroidal solenoid confines the magnetic field to its interior. To get the field inside the solenoid, choose the radius of the circular path such that it is inside the solenoid:



This time the enclosed current is the total number of turns in the solenoid N *multiplied by the current through the wire* I*, giving us the field inside:*

$$B\,(2\pi r) = \mu_o N I \quad \Rightarrow \quad B = \frac{\mu_o N I}{2\pi r}$$

Notice that this differs from the field in a straight solenoid in two ways: First, the field is not uniform within the solenoid – it is stronger closer to the center than it is near the outside of the solenoid. And second, it is the total number of turns (which is now finite) rather than the turn density that determines the field strength. Interestingly, the turn density for the toroid can be adjusted to "turns per radian" rather than turns per meter. This "angular turn density" is simply $\frac{N}{2\pi}$.

Local (Differential) Form of Ampére's Law

Just as with the case of Gauss's law, we can use a theorem from vector calculus to convert Ampére's law into local form. The name of the theorem in question is *Stoke's theorem*, and it states that for a vector field $\vec{B}\left(\vec{r}\right)$:

$$\oint \vec{B} \cdot \vec{dl} = \int \left(\vec{\nabla} \times \vec{B}\right) \cdot d\vec{A}\ , \tag{4.5.7}$$

where the closed path borders the area over which the area integral is performed. If we now apply Ampére's law to this and note that the current that passes through the area equals the flux of the current density, we get:

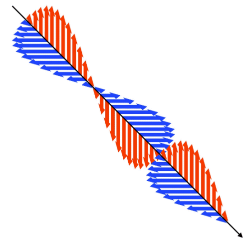
$$\int\left(\overrightarrow{\nabla}\times\overrightarrow{B}\right)\cdot d\overrightarrow{A}=\mu_o I_{encl}=\mu_o\int\overrightarrow{J}\cdot d\overrightarrow{A}\quad\Rightarrow\quad\overrightarrow{\nabla}\times\overrightarrow{B}=\mu_o\overrightarrow{J}\tag{4.5.8}$$

As with the local form of Gauss's law, this provides us with a means for doing calculations that would otherwise be more difficult using the integral form.

CHAPTER OVERVIEW

5: ELECTROMAGNETISM

The time has come to put our two disparate forces together. We know that they have electric charge in common as a source of field and a recipient of force, but here we will see that they have much more in common than this.

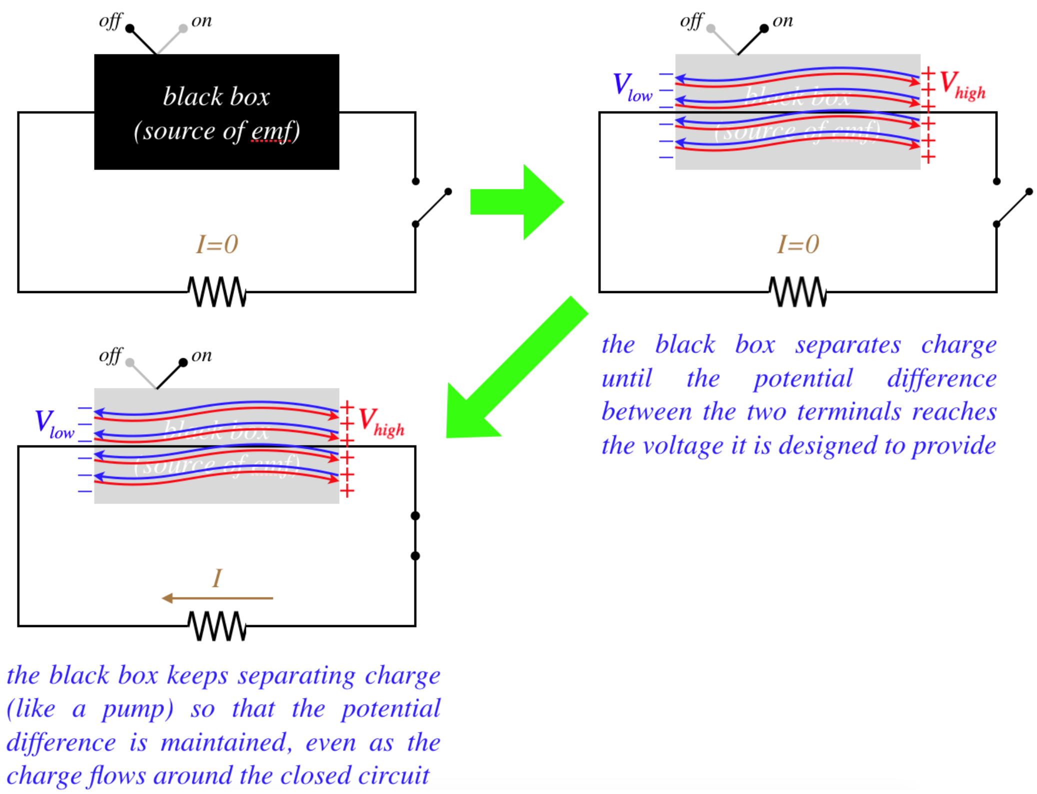


5.1: Magnetic Induction

Motional EMF

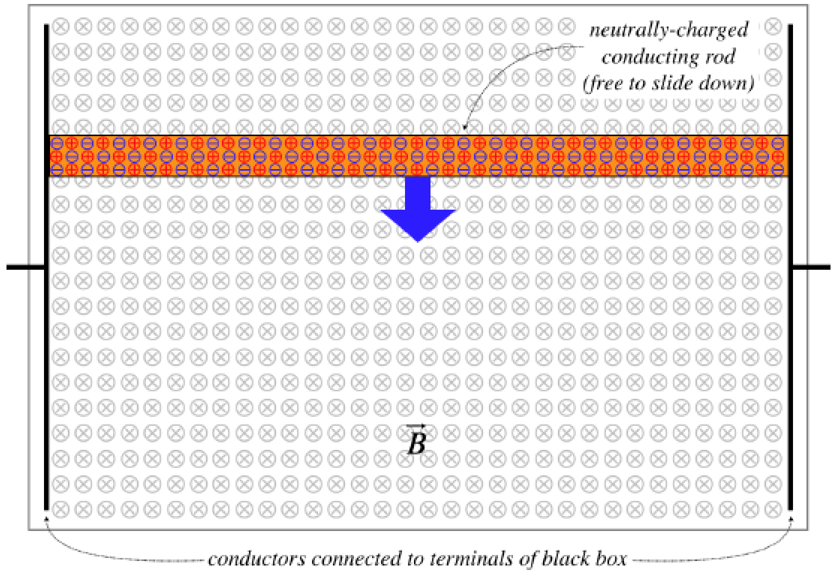
So far, our only encounter with emf has been through batteries, which separate positive and negative charge through a chemical reaction. We return now to a more general discussion of emf, so that we can apply it to what follows. To this end, let's consider an emf source that we'll describe as a black box.

Figure 5.1.1 – How Every Source of EMF Behaves



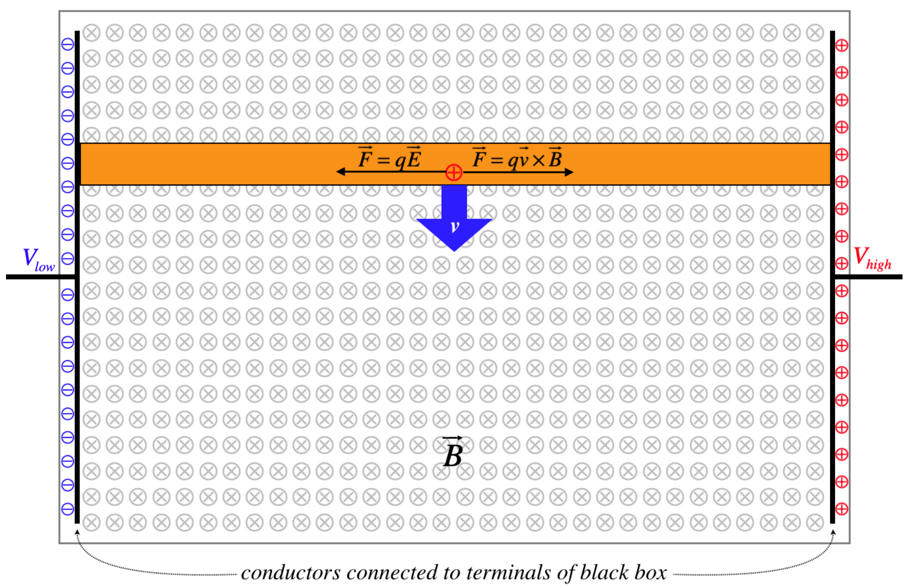
Given what we now know about magnetism, we can construct an emf source that doesn't depend upon the chemical reactions found in batteries. Suppose we replace the black box above with a pair of metal plates (across which the potential will be held at a fixed difference), with a conducting bar in contact with both that is free to slide. The other ingredient for this emf black box is a uniform magnetic field (say from a nearby bar magnet) in the region between the plates where the bar slides.

Figure 5.1.2 – New EMF Black Box



Now let's consider what happens as the bar slides down. The positive and negative charges in the bar are now moving through the magnetic field, and therefore experience magnetic forces in opposite directions. According to the right-hand-rule, positive charges will be pulled to the right, and negative charges to the left. They are confined to the conducting bar, so they cannot travel in circles – as free charges would – and end up collecting on the metal plates. They will continue to do so until the *electric* field of the separated charges is strong enough to oppose the magnetic forces on the charges.

Figure 5.1.3 – Motional EMF



The separated charge corresponds to a potential difference between the two plates. If the terminals of this black box are used to drive a current in a circuit, the magnetic force will continue to replace the charge as long as the bar continues moving. The amount of emf produced by this device (called *motional emf*) can be found by computing the potential difference from the electric field that balances the force on the charges by the magnetic field:

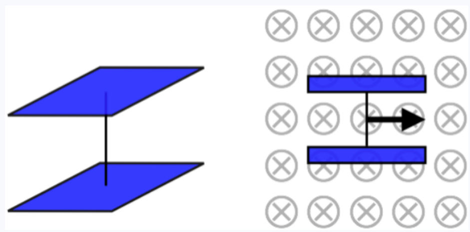
$$\mathcal{E}_{motional} = EL = vBL \;, \tag{5.1.1}$$

where L is the length of the conducting bar.

Clearly this stores energy in the black box, but where does this energy come from? When the charges start moving along the bar, they acquire a component of motion in the *horizontal* direction as well as the vertical component that has been forced upon them with the motion of the bar. The positive charge moving to the right feels a force upward by the magnetic field, but it can't leave the bar, so it pulls the bar upward (the negative charges moving in the opposite direction also pull the bar upward). This means that to keep the bar moving downward at a constant speed, a force needs to be applied from outside. This force does work on the bar (which doesn't gain kinetic energy, since it maintains a constant speed), and this work is what adds energy to the system.

Example 5.1.1

Two identical square conducting plates are oriented parallel to each other and are connected by a conducting wire as shown in the left diagram. This apparatus is then moved through a uniform magnetic field as shown in the right diagram (the thickness of the plates is negligible). The strength of the magnetic field is $1.5T$. The apparatus is moving at a speed of $8.0\frac{m}{s}$. Find the uniform charge density (including its sign) induced on the top plate.



Solution

Motional emf will create a potential difference between the two ends of the wire, which means there will be a potential difference between the two conducting plates. Once the steady potential difference is established, this essentially becomes a charged parallel-plate capacitor. Setting the motional emf equal to the potential difference between capacitor plates, gives:

$$\left.\begin{array}{l} \mathcal{E}_{motional} = Bvd \\ V_{capacitor} = \dfrac{Q}{C} \\ C_{parallel-plate} \end{array}\right\} \quad Bvd = \frac{Q}{C} = \frac{Qd}{A\epsilon_o}$$

The ratio $\frac{Q}{A}$ is the charge density, so we can immediately compute its magnitude:

$$\sigma = \frac{Q}{A} = \epsilon_o Bv = \left(8.85\times10^{-12}\frac{C^2}{N\cdot m^2}\right)(1.5T)\left(8.0\frac{m}{s}\right) = 1.1\times10^{-10}\frac{C}{m^2}$$

With the plates moving to the right and the magnetic field into the page, the RHR shows that positive charges in the wire will feel a force upward, which means the top plate will be positively charged.

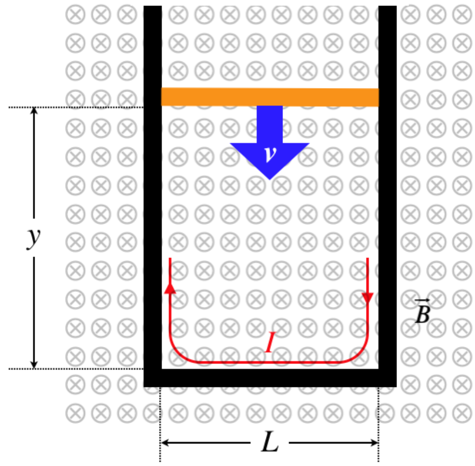
[It should be noted that one does not need to treat this as a parallel plate capacitor to achieve this answer. The charges stop flowing when the magnetic force balances the electric field force created by the separated charges. This electric field is, to a good approximation (the same one we make for parallel-plate capacitors), uniform between the plates, and at each plate the magnitude is $\frac{\sigma}{\epsilon_o}$. This achieves the same answer.]

Faraday's Law

Now that we know we can generate emf in conductors using magnetic fields, we will explore this phenomenon further. We'll start by looking at an alternate way of describing the phenomenon of motional emf that comes courtesy of a fellow named Michael Faraday, who did his work in this area in the middle of the 19th century.

Instead of stopping the analysis of the moving bar at the boundaries of the black box, let's include the entire circuit. When we do this, the magnetic field passes through a closed loop defined by the circuit, and we can write the magnetically-induced emf in terms of the *magnetic flux* Φ_B through that closed loop.

Figure 5.1.4 – Faraday's Model for Magnetically-Induced EMF



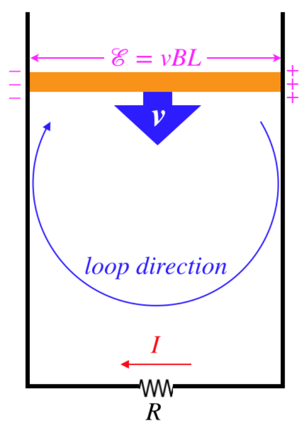
The emf induced in this physical system needs to be the same as it was above, and Faraday noted that the magnitude of the emf can be computed in terms of the time rate of change of the magnetic flux through the closed loop:

$$\Phi_B = \int \overrightarrow{B} \cdot d\overrightarrow{A} = BLy \quad \Rightarrow \quad |\mathcal{E}| = |vBL| = \left|-\frac{dy}{dt}BL\right| = \left|\frac{d}{dt}(BLy)\right| = \left|\frac{d\Phi_B}{dt}\right| \tag{5.1.2}$$

This may seem like a pointless rewriting of the motional emf result we got above, but in fact this turns out to be the more general rule for magnetically-induced emf. That is, there are more ways than just moving one of the sides of a rectangle to change the flux through a loop. Before we go into details of some of these variants, we need to take a moment to clarify what is meant by an emf being induced in a closed loop.

For the example given above, the potential difference created is obvious – the right side of the sliding bar is at a higher potential than the left side.

Figure 5.1.5 – Kirchhoff Loop Rule Applied to Sliding Bar Case



This means that if we employ Kirchhoff's loop rule to this circuit, we get:

$$+\mathcal{E} - IR = 0 \tag{5.1.3}$$

But Faraday's formulation of induced emf does not require a moving bar. For example, if the dimensions of the circuit don't change at all, the magnetic flux can still be changed by altering the magnetic field. If we try to do the same thing as above with Kirchhoff's loop rule, we don't have a specific segment of the loop where the potential jumps, so how do we account for the induced emf? Faraday's answer is that we simply modify Kirchhoff's loop rule – instead of having the sum of the voltage drops around the loop adding up to zero, they add up to the contribution of the changing magnetic flux:

$$sum\ of\ voltage\ drops\ around\ closed\ loop = \frac{d\Phi_B}{dt} = \frac{d}{dt}\int_{\substack{closed\\loop}} \overrightarrow{B} \cdot d\overrightarrow{A} \tag{5.1.4}$$

There is only one tricky detail left to clarify here – the sign convention. The area vector for the flux calculation is defined by the closed loop, but this vector needs a direction, and there are two possible directions to choose from. This direction is defined by our chosen loop direction and (what else?) the right-hand rule, in the same way that we defined the direction of the area vector for the magnetic moment; curl the fingers to trace the loop direction, and thumb points the direction of the area vector.

Let's try this for the case given above. The loop direction is clockwise, giving an area vector into the page. The magnetic field is also into the page, so the flux is a positive value. This flux is *decreasing with time*, which means that the time derivative of the flux is negative, giving for our modified Kirchhoff loop rule:

$$-IR = \frac{d}{dt}\int \overrightarrow{B}\cdot d\overrightarrow{A} = \frac{d}{dt}(+BLy) = Bx\frac{dy}{dt} = BL(-v) \tag{5.1.5}$$

If we had chosen the loop direction to be counterclockwise, then the voltage change across the resistor would be positive (because the labeled direction of the current is opposite to the loop direction), and the flux would be negative, giving the same final result. Typically this is expressed in terms of the emf that needs to be added into the Kirchhoff loop to again give a zero loop sum. That is:

$$sum\ of\ voltage\ drops\ around\ closed\ loop + \mathcal{E}_{induced} = 0 \quad \Rightarrow \quad \mathcal{E}_{induced} = -\frac{d\Phi_B}{dt} \tag{5.1.6}$$

There is one more detail that needs to be added here. If the circuit consists of a coil of wire with many (N) turns in it, then the flux through every turn contributes to the induced emf, giving the final form of what is known as *Faraday's law*:

$$\mathcal{E}_{induced} = -N\frac{d\Phi_B}{dt} \tag{5.1.7}$$

Lenz's Law

Keeping track of the sign convention in Ampére's law can be an odious task, but it turns out there is an easier way to determine the direction that the induced emf would seek to drive a current. Not only is it easier, but it provides some useful physical insight into *why* the minus sign appears in Faraday's law.

Suppose we have a closed loop of conducting wire in the plane of this page, through which we suddenly introduce a magnetic field into the page. This change in magnetic flux will induce an emf in the loop, which will in turn result in a current flowing through the loop. Which way will this current flow? A Russian physicist named Emil Lenz argued that the current would have to be induced *counter*clockwise, for the following reason: The induced current will contribute to the total magnetic field passing through the loop (magnetic dipoles also have fields!), and if the induced current was clockwise, then the additional field from the current in the wire would *increase the flux* that has already been increased. This would induce more current in the same direction, which would induce more current, and so on. Obviously the energy in the circuit would then grow without bound, which isn't possible. He therefore landed upon what is now known as *Lenz's Law*:

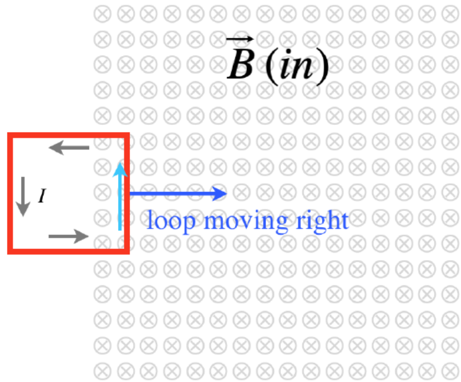
The emf induced in a circuit will be such that a current resulting from it will produce a secondary magnetic flux that seeks to undo the change in flux that induced the emf in the first place.

It should be noted that induced emf's don't always drive a current (e.g. there may be a battery present, and the induced emf may only slow the current), but it is not the resulting current that matters – the induced emf will "try" to induce a current in a direction that counters the change in flux. Also, the amount of current induced generally will not entirely undo the change in flux – this law only provides the *direction* of the induced emf.

An Illustrative Example

It is useful to look at a few concrete examples of magnetic induction. The first involves a closed conducting loop moving through a region of uniform magnetic field. In this case, we can view it either in terms of Faraday's law, or in terms of motional emf.

Figure 5.1.6 – Conducting Loop Enters Uniform Magnetic Field



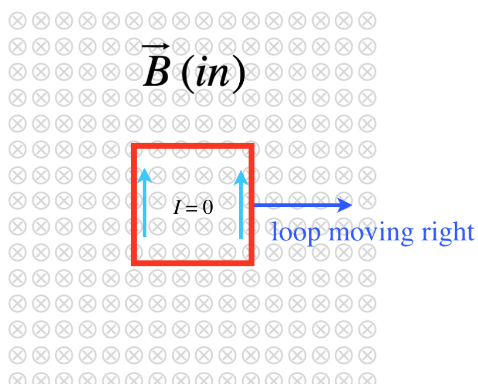
Motional EMF Explanation

When the loop first enters the field, only the right side of the loop is moving through the field, so the motional emf creates a higher potential at the top of this side than at the bottom. The imbalance drives a current counterclockwise.

Faraday/Lenz Explanation

As the loop enters the field, the magnetic flux through it increases into the page. This induces an emf that seeks to drive a current whose magnetic field will oppose this change. To reduce the flux inward, a field needs to be created that points out of the page inside the loop. From the right-hand-rule, this requires a counterclockwise currrent.

Figure 5.1.7 – Conducting Loop Stays Within the Uniform Magnetic Field



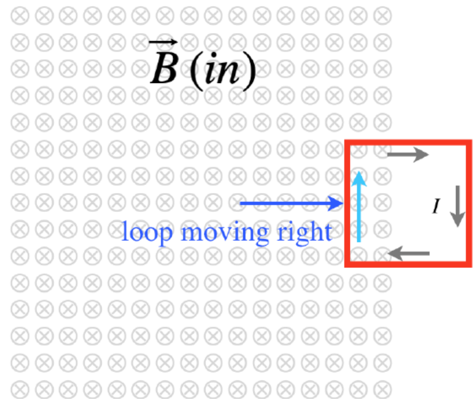
Motional EMF Explanation

With the loop entirely within the field, *both* the left and right sides of the loop exhibit motional emf, holding the top segment of the loop at a higher potential than the bottom segment. But just as if two identical batteries were placed in both sides of the circuit with their positive terminals facing up, this does not result in any current around the loop.

Faraday/Lenz Explanation

As the loop remains within the uniform field, the flux through the loop doesn't change, because none of the factors that goes into the calculation of the flux changes. With an unchanging flux, there is no induced emf in the loop, and no current as a result.

Figure 5.1.8 – Conducting Loop Stays Within the Uniform Magnetic Field



Motional EMF Explanation

Once again the loop has one of its sides outside the field, so only one side exhibits motional emf. This again results in an imbalance that drives a current. In this case, the current is in the opposite direction as when the loop entered the field.

Faraday/Lenz Explanation

The loop is losing inward flux with time as it leaves the field, so the emf induced will act to create a current that opposes this loss of flux. The direction of the current that will bolster the field inside the loop into the page is clockwise.

Not withstanding this enlightening example, we can't handle every case by resorting to a motional emf argument. For emf to be induced, we only need that the flux through a closed loop changes. Given the definition of the flux, there are many ways for this to happen:

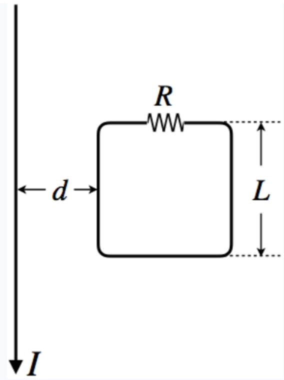
$$\Phi_B = \int \vec{B} \cdot d\vec{A} = \int B dA \cos\theta \tag{5.1.8}$$

- area changing with time (e.g. sliding bar example earlier)
- magnetic field changing with time (e.g. move a magnet farther or closer to change the field strength)
- change the angle θ between the field and the area (e.g. a rotating loop in a uniform field)
- change the the integral (e.g. move the loop between regions where the magnetic field is different).
- any combination of the above

When we calculate the flux, we need to add up the small flux increments over the whole surface… But what surface? While we tend to think of the surface as being the flat area defined by the loop, there are actually infinitely-many surfaces defined by the loop as a border. Well, it turns out that all of the surfaces result in the same flux, for the following reason: Consider two different surfaces at once: The flux that goes through one, must go out of the other one (since they are bordered by the same closed loop), because as we know from Gauss's Law for magnetism, the magnetic flux for any closed surface is zero (no magnetic monopoles). Therefore the flux through every surface defined by the boundary is the same.

Example 5.1.2

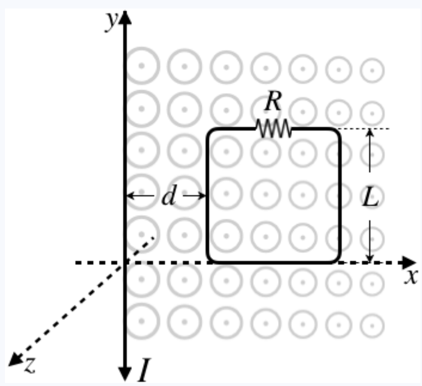
A square loop of wire with a total resistance R has a side-length of L. It resides near a long, straight wire such that the long wire is in the same plane as the loop, with one of its sides parallel to the wire a distance d from it, as shown in the diagram.



a. *Compute the magnetic flux through the loop due to a current I flowing in the direction shown in the diagram.*
b. *If the current in the long wire is changing, then the magnetic field flux through the loop will also be changing. Suppose that the current in the long wire is increasing at a constant rate of $\frac{dI}{dt}=\alpha$. Find the current induced in the loop, including its direction.*

Solution

a. We know the magnetic field for a long-straight wire gets weaker at greater distances, which means we have to perform the flux integral. The diagram below shows the magnetic field and introduces some axes for performing the math required.



The magnetic field strength in the x-y plane in terms of the distance x from the long wire is given by:

$$B=\frac{\mu_o I}{2\pi x}$$

We take as a differential area element a thin vertical slice down the length of the circuit. This slice has an area of $dA=Ldx$, and the field is constant throughout the slice. Also, the area vector is parallel to the magnetic field, so choosing the loop direction as counterclockwise, the angle between the field and the area vector is 0^o. The flux integral is therefore:

$$\Phi_B=\int \overrightarrow{B}\cdot d\overrightarrow{A}=\int B dA\cos\theta=\int\limits_{x=d}^{x=d+L}\left(\frac{\mu_o I}{2\pi x}\right)Ldx\cos 0^o=\frac{\mu_o IL}{2\pi}\ln\left[\frac{d+L}{d}\right]$$

b. The emf induced in the loop (which has only one turn in it) is found using Faraday's law:

$$\mathcal{E}_{induced}=-\frac{d\Phi_B}{dt}=-\frac{\mu_o L}{2\pi}\ln\left[\frac{d+L}{d}\right]\frac{dI}{dt}=-\frac{\mu_o L\alpha}{2\pi}\ln\left[\frac{d+L}{d}\right]$$

The magnitude of the current is this emf divided by the resistance:

$$\frac{\mu_o L\alpha}{2\pi R}\ln\left[\frac{d+L}{d}\right]$$

The direction of the current will be such that it provides a field that will seek to counter the change. The current that is causing the field is increasing, so the flux out of the page in increasing. The induced current will therefore produce a magnetic field inside the loop that points into the page, which means it must flow clockwise.

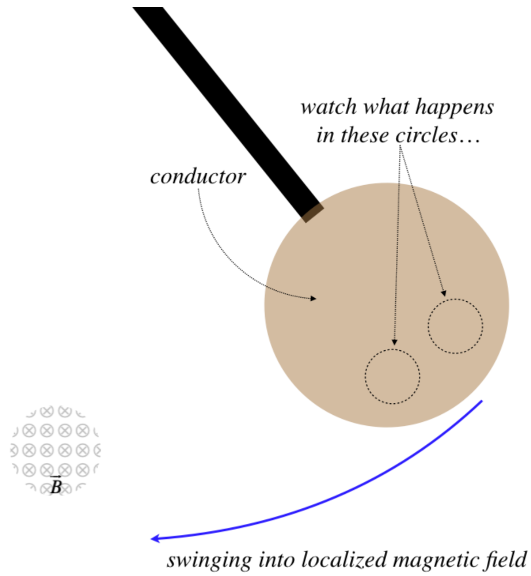
5.2: Consequences and Applications of Induction

Eddy Currents and Magnetic Damping

In our original discussion of motional emf, we pointed out that the bar actually has to be *pushed* through the magnetic field in order to maintain a constant speed, since the induced current experiences a force that opposes the motion. We noted that this was necessary for energy to be conserved. This same effect can be exploited as a breaking device.

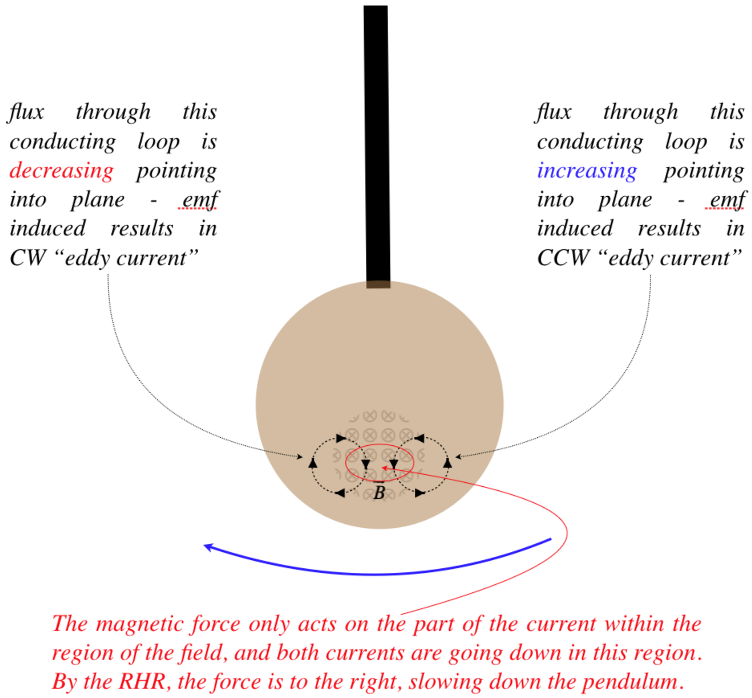
Consider a conducting plate moving through a localized magnetic field (say past the pole of a magnet).

Figure 5.2.1a – A Conductor Swings into a Localized Magnetic Field



When this conductor reaches the vertical and moving to the left through the magnetic field, the flux through each of these circles is changing in different ways.

Figure 5.2.1b – A Conductor Swings into a Localized Magnetic Field



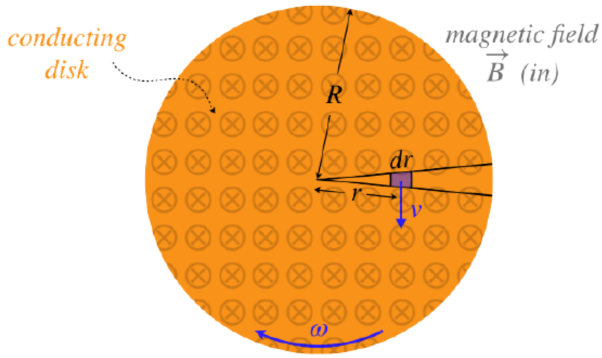
As usual, we find that when currents are induced, the energy they require is removed from the mechanical system inducing that current. These swirly *eddy currents* that form in the conductor due to magnetic induction (they appear everywhere in the conductor that passes through the field, not just the two spots shown in the figure) also require energy, and as we can see, the magnetic force on these eddy currents acts to damp the motion of the conductor through the field. Ultimately the energy that forms these currents is turned into thermal energy as the current passes through the resistance in the conductor.

Electric Generators

The natural next step to take in magnetic induction is to produce consistent, usable, electrical current. This is done through the use of a device once referred to as a *dynamo*, and now is generally known as a *generator*. These come in lots of varieties, but they all ultimately convert mechanical energy (which can come from the potential energy of water falling through a turbine in a dam, the kinetic energy of wind turning a windmill, or a hamster in an exercise wheel) into emf that can drive a current. We will focus here on a couple of designs that result in direct current, though nowadays generators that produce an oscillating *alternating current* are far more common.

The very first design for such a dynamo was by Faraday himself, and it is essentially a clever way of using the "sliding bar" method of generating motional emf (described in Section 5.1) without the problem of the bar reaching the end of its track. If we fix one end of the bar to a pivot, and spin it in a circle whose plane is perpendicular to a uniform magnetic field, then it can constantly produce a motional emf between its two ends. But Faraday took this one step further: Imagine gluing many thin pie-slice wedges of these bars together so that they form a solid conducting disk, then spin the disk in the field. Then a motional emf will be generated between the center of the disk and its edge. One lead of a circuit can be connected to the center, and the other to a conducting brush that makes contact with the outer edge, and as long as one keeps the disk rotating, current will flow through the circuit.

Figure 5.2.2 – A Faraday Disk Dynamo



To compute the emf generated between the center and the rim, we consider the small section of conductor (that is the "metal bar" moving through the field). Motional emf is created between the two ends of this segment, with the right-hand-rule telling us that the side farther from the center of the disk will have the higher potential. The length of the bar is dr, so the tiny emf generated is, according to Equation 5.1.1:

$$d\mathcal{E} = v\ B\ dr \tag{5.2.1}$$

The segment is traveling in a circle, so its speed v is related to its distance from the center and the disk's rotational speed by $v = r\omega$. The same potential difference exists for every segment of conductor in the pie slice, so these can be added together to get a total potential difference between the center of the disk and its rim. Noting that both $omega$ and B are assumed to be constant, we get:

$$\mathcal{E}(\text{center to rim}) = \int\limits_{r=0}^{r=R} (r\omega)\, B\ dr = \frac{1}{2}R^2\omega B \tag{5.2.2}$$

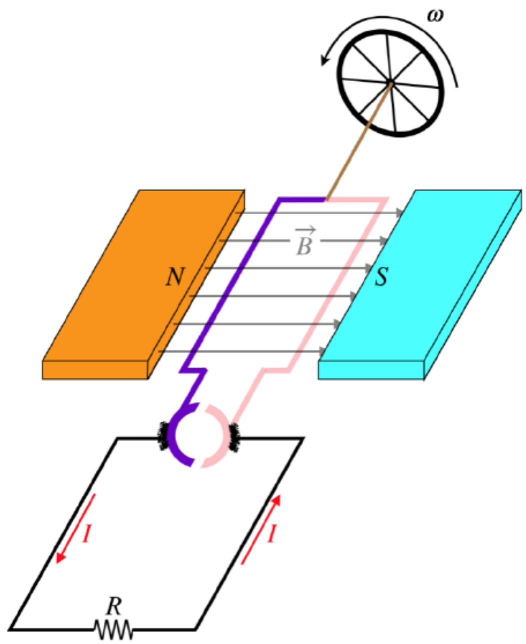
A couple of notes to make here:

- Every piece slice is doing this at the same time, making the entire rim at the same higher potential.
- In the case above, if the dynamo is spun in the opposite direction, or if the field points in the opposite direction, then it would be the center of the disk that would be at the higher potential.
- As always, if a current is flowing thanks to this device, the energy is coming from whatever mechanical energy is keeping the disk turning, which means that there must be a torque opposing the rotation of the disk as the current flows (against which work is being done). If this torque serves to slow the rotation while the current charges a battery, we have all the essential ingredients necessary for *regenerative braking* currently common in electric and hybrid automobiles.

A nice feature of the disk dynamo is that as long as the rotational speed is held constant, the potential difference is unchanged. The next generator we will look at is a somewhat less cumbersome design than the disk dynamo, but while it does provide a direct current (i.e. a current that doesn't change direction), the emf is not a steady one.

We have seen this other design before – it is identical to that of the DC motor!

Figure 5.2.3 – A Reversed DC Motor is a DC Generator



With the wheel being turned at a constant rate, we can use Faraday's law to compute what the induced emf is around the closed circuit as a function of time (we'll compute the absolute value of the emf, and figure out the direction with Lenz's law afterward):

$$|\mathcal{E}| = \left|\frac{d}{dt}\Phi_B\right| = \left|\frac{d}{dt}[BA\cos\theta]\right| = \left|BA\frac{d}{dt}\cos\theta\right| = \left|BA\left[-\sin\theta\frac{d\theta}{dt}\right]\right| = \omega BA\,|\sin\theta| \tag{5.2.3}$$

At the moment depicted in the diagram, the emf is a maximum. We can deterine the direction of the induced current by considering what happens to the loop an instant later – the pink side is rising, so the flux is increasing left-to-right through the loop with the pink side above the purple side. The induced current will see to reduce this increase in flux, which from the right-hand-rule tells us that the emf is induced such that the current will flow up the pink side and down the purple side, giving the direction indicated in the figure.

It is important to note that unlike the disk dynamo, this DC generator does not produce a constant emf (it varies according to $\sin\theta$). The presence of the commutator keeps this current always going in the same direction, but it varies from one moment to the next. Many generators like this do not include a commutator, producing alternating current instead. Which generator is used largely depends upon the application. For example, if you want to charge a battery or a capacitor, you don't want the emf constantly changing direction (causing the charge to go into and out of the device you are charging), but if you are turning on a heating element, the resistor will produce thermal energy no matter which way the current is going.

Finally, we will note (as we always do) the fight that goes on between the electrical and mechanical sides of this device. When we are running a motor, the coil is turning in the magnetic field and generating an emf opposite to the direction of the applied emf, reducing the current that goes into driving the motor (this is called *back emf*). Similarly, when running as a generator, the current generated is in a magnetic field, resulting in a torque that opposes the rotation. This assures that we have to constantly do work to keep the generator running, converting the work added into electrical energy that then accomplishes whatever it is doing.

Diamagnetism

In Section 4.4 we discussed ferromagnetism and paramagnetism, two phenomena that result in materials acting as sources of magnetic fields. The former is responsible for permanent magnets, and the latter for the ability of magnets to stick to otherwise unmagnetized objects like paper clips and refrigerators. These phenomena occurred because dipoles that occur naturally in matter are allowed to align with the external fields.

But another effect also occurs when a magnetic field is introduced to materials. The change of the magnetic flux through the material induces an emf that results in the electrons' average motions opposing the flux according to Lenz's law. This occurs in everything (even paramagnetic substances), but the effect is much smaller than paramagnetism, so it is much harder to witness, and it doesn't "undo" para- or ferro- magnetism. Though it is weak, it can be seen in some substances (most notably, water) under the influence of very strong magnetic fields.

There is one way that we can very dramatically witness diamagnetism in action. This involves a substance that allows for a great deal of induced current to be created at the atomic level, due to the very low resistance of the substance. This substance is known as a *superconductor*, because its resistance is effectively zero. Let's imagine what would happen if we set a magnet down on a slab of superconductor. The induced current would oppose the new flux, which means that the magnetic field from the superconductor will be in the opposite direction of the applied field, causing a repulsive force, and levitation!

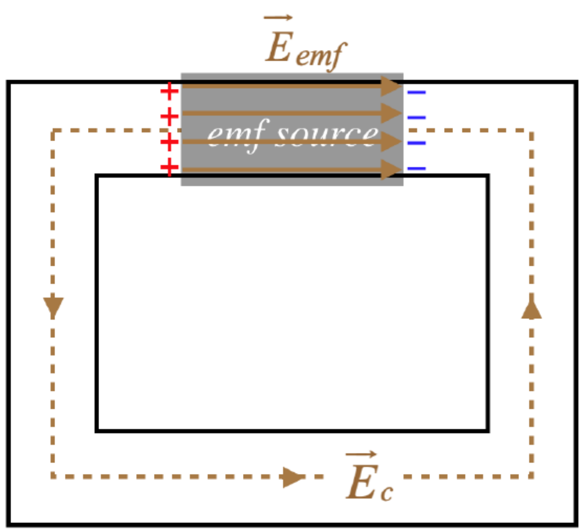
It turns out that this doesn't come close to describing the full story of magnetic effects on superconductors. There is much more that is even stranger (something called the *Meissner effect*), but a treatment of this topic involves quantum mechanics, and is beyond the scope of this class.

Induced Electric Fields

When we talked about motional emf and moved on to Faraday's law, we glossed-over the fact that for motional emf we could explain the induced current in terms of forces on particles, but for Faraday's law we didn't have a similar explanation. For example, suppose we place a solenoid inside a loop and vary the current. The magnetic field outside the solenoid remains zero, which means it cannot be exerting forces on the charges in the conductor, so how does this induce a current? To answer this, we need to return to where the force on charges in a conductor (causing current) comes from.

If we introduce a battery to a closed circuit, there are really two electric fields involved. One of these is *within* the battery (produced by the chemical reactions), which acts to separate the free charge inside the battery to the two terminals. The second electric field is the one that is produced by the battery terminals in the conductor through which the electric current flows. If we add up the voltage drops around the full circuit, both of these fields contribute.

Figure 5.2.4 – Voltage Drops Around a Circuit with a Battery

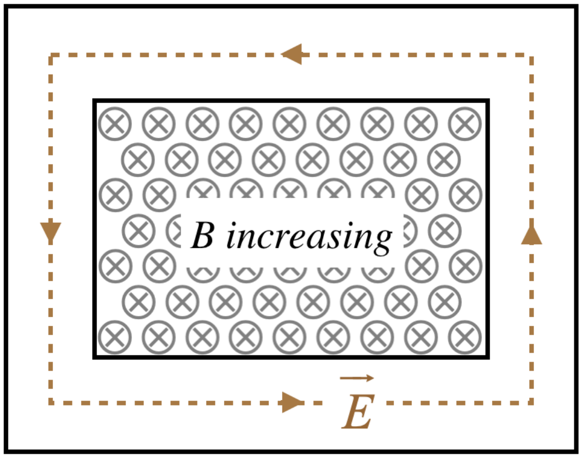


Integrating the electric field around the closed path that is the circuit gives the sum of the voltage drops, and whichever direction we choose for our closed path, one of the electric fields is along the path direction and one is opposite to it. These two contributions cancel to give the usual Kirchhoff loop rule result:

$$\oint \overrightarrow{E}\cdot\overrightarrow{dl} = \int\limits_{conductor} \overrightarrow{E}_c\cdot\overrightarrow{dl} + \int\limits_{emf\ source} \overrightarrow{E}_{emf}\cdot\overrightarrow{dl} = 0 \tag{5.2.4}$$

But when the source of the current is a changing magnetic flux due to a changing magnetic field strength through a circuit of fixed dimensions, the result is different. There is only *one* electric field present in this case – the one that drives the current. There is no section of the circuit where an emf source has an opposing field.

Figure 5.2.5 – Voltage Drops Around a Circuit Driven by a Changing Magnetic Field



In this case, when we take the same integral around a closed path, we don't have two sections that cancel, and the result is a non-zero integral:

$$\oint \vec{E} \cdot \vec{dl} \neq 0 \tag{5.2.5}$$

This is agreement with Faraday's "modified Kirchhoff loop rule" explanation, given in Equation 5.1.4. Instead of talking about the changing flux inducing an emf, we can directly relate the magnetic and electric fields in the following way:

A changing magnetic field induces an electric field that circles around it in the direction defined by Lenz's law.

We can now use Equation 5.1.4 to specify this observation mathematically:

$$\text{sum of voltage drops around closed loop} = -\oint \vec{E} \cdot \vec{dl} = \frac{d}{dt} \int \vec{B} \cdot d\vec{A} \tag{5.2.6}$$

We can turn this integral equation into a differential equation with the help of Stokes' theorem:

$$\oint \vec{E} \cdot \vec{dl} = \int \left(\vec{\nabla} \times \vec{E}\right) \cdot d\vec{A} \quad \Rightarrow \quad \vec{\nabla} \times \vec{E} = -\frac{d}{dt}\vec{B} \tag{5.2.7}$$

This is *Faraday's law in local (or differential) form*.

Back when we studied electrostatics, we noted that the electric field could be written as the negative gradient of the electrostatic potential, and that this meant that the curl of the electric field vanishes (Equation 2.2.13). This is clearly not the case when there is a changing magnetic field nearby, which means that an electric field in this circumstance *cannot be written as a negative gradient of a potential field*. Indeed, by now we have left electrostatics far behind (we now consider currents flowing all the time), so it isn't surprising that we would lose this special relation.

One last comment here: An electric charge in space (not in a conductor with resistance) that follows a path that traces the electric field that circulates around a changing magnetic field, *will be going faster or slower* (depending upon the direction relative to the electric field and the sign of the charge) when it gets back to where it started. That is, this is not a conservative force! Not to worry, energy is still conserved overall, because the accelerated charge is not a closed system. Magnetic fields are produced by other moving charges, and for the magnetic field to change, their motions need to change. Together this interplay forms a very complex dance, and careful accounting of all the energy shows that it does remain constant for the whole system.

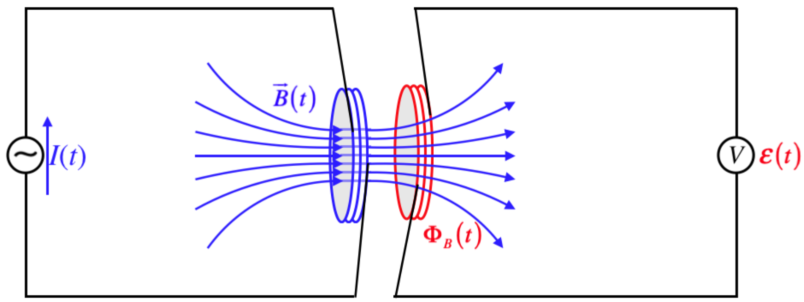
5.3: Inductance

Mutual Inductance

In keeping with our tradition of following the lead from electricity when discussing magnetism (Coulomb vs. Biot-Savart, dipole math, Gauss vs. Ampére, etc.), it's time we tackled magnetism's version of capacitance. Just to review, capacitance was purely a static charge notion – a way of storing energy on two conductors that hold opposite charges. Magnetism is anything but a static charge phenomenon, so its version of capacitance will have to be quite different. It is different, but there are many parallels as well.

If we place two circuits near each other and change the current flowing through one of them, then the magnetic field for that changing current will also change. If this magnetic field results in a flux through the second circuit, then its change can induce a current in the second circuit. So somehow we can transfer energy from one circuit to another without the circuits even being connected to each other. One way to explain this is to assume that there is energy present in the magnetic field itself. We already know that energy is contained in an electric field, so this is not a surprising revelation. Here's a diagram of this physical situation:

Figure 5.3.1 – Mutual Inductance



We have introduced a new symbol here. The circle with the wavy line indicates a variable current source (this explains why there is the time-dependent current variable right next to it). This variable current $I(t)$ passes through the left coil, resulting in a time-dependent magnetic field $\overrightarrow{B}(t)$. This field passes through the right coil, resulting in a time-varying flux $\Phi_B(t)$. Faraday's law says that this variable flux results in an induced emf $\mathcal{E}(t)$ registered in the voltmeter.

Our goal here is to mathematically relate the emf developed in the secondary coil to the variable current sent through the primary coil. We start by noting that the magnetic field strength is directly proportional to the current through the *primary coil* (the one on the left in the figure above), and the magnetic flux through the *secondary coil* is proportional to the magnetic field strength. Both of these facts are based on the idea that the field and its relation to the area through which it passes has the same shape, given that the geometry of the construction of this device remains unchanged. This should sound very familiar. Putting these proportionalities together, we get that the flux through the secondary coil (which we'll say has N turns) is proportional to the current through the primary coil, and we can define a proportionality constant M, called the *mutual inductance* of the construction:

$$N\Phi_B(t) = MI(t) \tag{5.3.1}$$

This gives us the relationship between the input current and the resulting emf that we were looking for, from Faraday's law:

$$\mathcal{E}(t) = -M\frac{d}{dt}I(t) \tag{5.3.2}$$

This is the magnetic analog to the electrostatic $Q = CV$. Like the case of capacitance, the mutual inductance is only a function of how the device is constructed – it doesn't change when the current through it is changed, just as capacitance is not changed by altering the amount of charge on the plates. The units for inductance are:

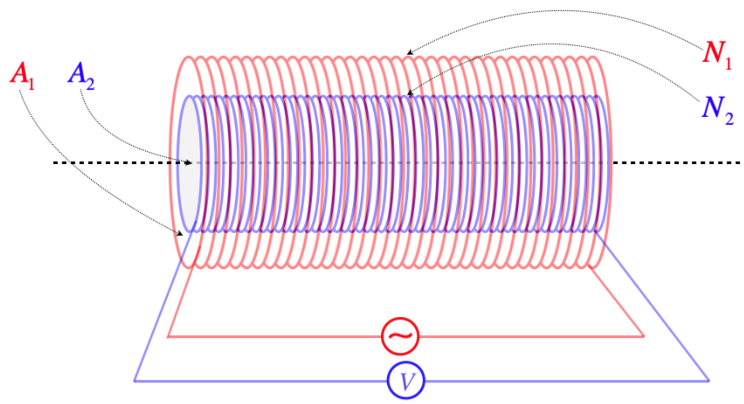
$$[M] = \left[\frac{B\cdot A}{I}\right] = \frac{T\cdot m^2}{A} \equiv H \quad (\text{"henry"})$$

One thing that seems to distinguish mutual inductance from capacitance is the apparent one-way nature of it – there are distinct primary and secondary coils, while a capacitor just has two conductors, either of which can take the positive charge. Well, it turns out that mutual inductance is equally reversible! The reason is that the number of turns in the primary coil is proportional to the magnetic field strength in the same way that the number of turns in the secondary coil is proportional to the total flux. That is, if we swap the current source and the voltmeter in the figure above, we get the same result. We can express this mathematically by labelling everything (number of turns, flux, and current) for each coil differently:

$$M=\frac{N_2\Phi_2}{I_1}=\frac{N_1\Phi_1}{I_2} \tag{5.3.3}$$

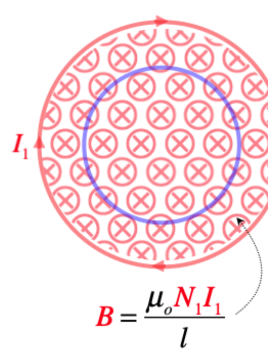
Let's see that this is true (as well as explore some nuances of the physics) by looking at an example. The physical setup consists of two coaxial solenoids of equal lengths, but their own cross-sectional areas and number of turns.

Figure 5.3.2 – Mutual Inductance of Coaxial Solenoids – Outer Cylinder Carries Current



In this setup, we have the applied varying current going through the outside solenoid, and the changing flux within the inside solenoid. The outer solenoid's field is uniform throughout its interior, and only the portion within the region of the inner solenoid contributes to the flux.

Figure 5.3.3 – Flux Through Inner Solenoid of Field from Outer Solenoid

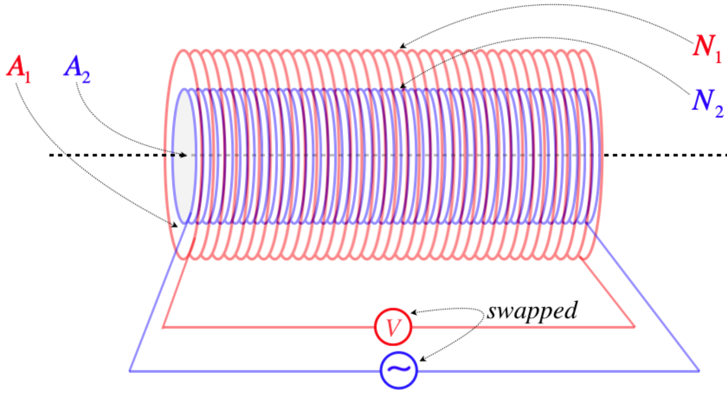


The field produced by the outer solenoid is easily calculable from the current, the number of turns, and the length, and the flux is easy to calculate from there:

$$\Phi_2=BA_2=\left(\frac{\mu_oN_1I_1}{l}\right)A_2 \quad\Rightarrow\quad M=\frac{N_2\Phi_2}{I_1}=\frac{\mu_oN_1N_2A_2}{l} \tag{5.3.4}$$

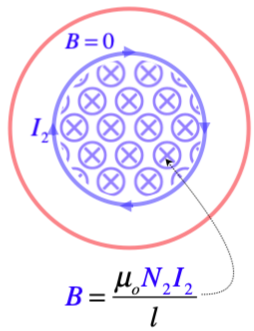
We can see that this result depends only upon the structure of the two solenoids. Okay, so let's check to see if the reversibility property holds. We now put the varying current through the inner solenoid.

Figure 5.3.4 – Mutual Inductance of Coaxial Solenoids – Inner Cylinder Carries Current



As before, we need to determine the flux through the secondary coil due to the current in the primary, but this time the picture is slightly different.

Figure 5.3.5 – Flux Through Outer Solenoid of Field from Inner Solenoid



The flux through the outer solenoid comes in two parts: The flux through the central region due to the field inside the inner solenoid, and zero flux through the outer region:

$$\Phi_1 = B_{inner} A_2 + B_{outer} \left(A_1 - A_2\right) = \left(\frac{\mu_o N_2 I_2}{l}\right) A_2 + 0 = \frac{\mu_o N_2 I_2 A_2}{l} \tag{5.3.5}$$

Now plug this into the definition of mutual inductance and find that it is the same as when the roles of the solenoids were reversed:

$$M = \frac{N_1 \Phi_1}{I_2} = \frac{\mu_o N_1 N_2 A_2}{l} \tag{5.3.6}$$

Self-Inductance

We can now see how one circuit can affect another without being in physical contact, but in fact there really doesn't need to be two coils in order to witness an effect of magnetic induction. Suppose we have a varying current in a circuit containing a single solenoid. As the current in the solenoid varies, the magnetic field inside the solenoid changes, which in turn changes the magnetic flux through that same solenoid. That induces an emf across the solenoid, which will in turn have an effect on the current through it! This process is known as *self-inductance*.

We actually define self-inductance in the same way that we defined mutual inductance – the ratio of the total flux through the N coils to the current that supplies the magnetic field. Naturally the units are therefore the same as mutual inductance.

$$L \equiv \frac{N\Phi}{I} \tag{5.3.7}$$

In practical terms, the only reasonable device for which we can compute the self-inductance is a solenoid, as its uniform magnetic field makes the flux easy to compute:

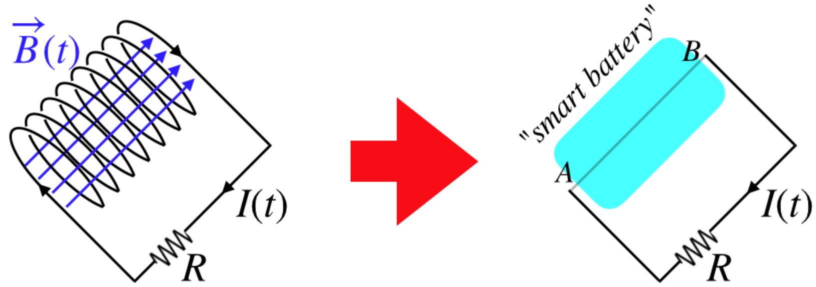
$$L_{solenoid} = \frac{N\Phi}{I} = \frac{N\left(BA\right)}{I} = \frac{N\left(\frac{\mu_o NI}{l}\right)A}{I} = \frac{\mu_o N^2 A}{l} \tag{5.3.8}$$

As with mutual inductance, this value L is a constant that depends only upon the structure of the solenoid (hereafter referred to as an *inductor*). The difference here of course is that the emf induced by the current/field/flux change acts to affect the same current that causes the flux change to begin with. In particular:

$$N\Phi = LI \quad \Rightarrow \quad N\frac{d}{dt}\Phi = \frac{d}{dt}(LI) \quad \Rightarrow \quad \mathcal{E} = -N\frac{d}{dt}\Phi = -L\frac{d}{dt}I \tag{5.3.9}$$

How do we interpret this? Think back to what is happening here – Faraday's law is in play, and this means that an emf is induced (like a battery, but without the chemical reaction). So we can think of this as a "smart battery" that adjusts its emf (magnitude and direction) according to what is happening to the current. The minus sign indicates that this smart battery adjusts itself to oppose changes. If the current is dying down, the inductor comes to the rescue and provides an emf to try to help maintain the current, and if the current is growing, then the inductor provides an emf to try to slow the increase.

Figure 5.3.6 – Inductor Behaves Like a "Smart Battery"



In the figure above, if the current is dying-down, the magnitude of the magnetic field diminishes, reducing the flux through the inductor. The induced emf across the inductor will be such that it will seek to bolster the current through itself (which also goes through the resistor). This requires that the potential at terminal B of the smart battery be greater than the potential at terminal A. If the current is increasing, then to fight this change, the potential of terminal A of the smart battery must be greater than the potential at terminal B.

Let's use Kirchhoff's loop rule on this circuit. Choosing a loop direction of clockwise, we have:

$$V_B - V_A - IR = 0 \tag{5.3.10}$$

If the current is decreasing, then its derivative is negative. In this case, the inductor acts like a smart battery with its positive terminal at position B, so the voltage drop across the inductor for our clockwise loop needs to be positive. If the current is instead increasing, then the derivative of the current is positive, and the smart battery reverses direction, making the voltage drop across the inductor for the clockwise loop negative. Both of these scenarios are in agreement with the conclusion above that the potential drop across an inductor satisfies $\mathcal{E} = -L\frac{dI}{dt}$.

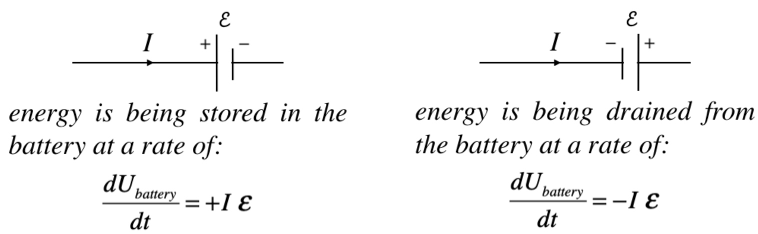
5.4: Inductors in Circuits

Magnetic Field Energy

Inductors are what we were looking for – a device that goes into a circuit like a capacitor which involves magnetic rather than electric fields. Several chapters ago, we said that the primary purpose of a capacitor is to store energy in the electric field between the plates, so to follow our parallel course, the inductor must store energy in its magnetic field. We can calculate exactly how much is stored using tools we already have.

Suppose we start building up a current from zero into an inductor. With no current in it, there is no magnetic field and therefore zero energy, but as the current rises, the magnetic field grows, and the energy stored grows with it. We actually have a way of determining the *rate* at which the energy stored is growing from what we know already. There is no resistance to worry about here, so none of the energy is lost to thermal, which means that we can write the power as the product of the current and the voltage difference. As a reminder, power delivered to or by a battery is plus-or-minus the product of the current and the emf of the battery:

Figure 5.4.1 – Power Charging or Discharging a Battery



With the idea of an inductor behaving like a smart battery, we have method of determining the rate at which energy is accumulated within (or drained from) the magnetic field within the inductor. If the positive lead of our smart battery is facing the incoming current, it must be because the current is increasing. This results in an *increase* in the energy stored in the inductor, and sure enough, an increase in current corresponds to an increase in the magnetic field strength within the inductor. The reverse argument for an inductor where the current (and therefore field) is decreasing also fits perfectly. The math works easily by replacing the emf of the battery with that of an inductor:

$$\frac{dU_{inductor}}{dt} = I\left(L\frac{dI}{dt}\right) = LI\frac{dI}{dt} \tag{5.4.1}$$

We can now determine the energy within the inductor by integrating this power over time:

$$U_{inductor} = \int P dt = \int \left(LI\frac{dI}{dt}\right) dt = L\int IdI = \frac{1}{2}LI^2 \tag{5.4.2}$$

There is clearly a resemblance of this energy to that of a charged capacitor, though the parallels are not immediately obvious. It seems reasonable to relate the charge to the current, because in each case, these are what is accumulated within the device. This would mean that the parallel between capacitance and self-inductance is $C \leftrightarrow L^{-1}$. This parallel only goes so far, however. For example, it doesn't work for $Q = CV$. For energy considerations, however, it does work well, and we will see that this extends to field energy.

The potential energy stored within a solenoid (which, as we stated above, is pretty much the design of every inductor) can be written in terms of the magnetic field within. For this we need the self-inductance of a solenoid (Equation 5.3.8), and the field of a solenoid (Equation 4.4.13):

$$U_{solenoid} = \frac{1}{2}LI^2 = \frac{1}{2}\left(\frac{\mu_o N^2 A}{l}\right) I^2 = \frac{1}{2\mu_o}\left(\frac{\mu_o NI}{l}\right)^2 (A\cdot l) = \frac{1}{2\mu_o}B^2\,(A\cdot l) \tag{5.4.3}$$

The quantity $A\cdot l$ is the volume of the solenoid, so dividing both sides by this gives the energy density of the magnetic field within the solenoid:

$$u_{solenoid} = \frac{U_{solenoid}}{volume} = \frac{1}{2\mu_o}B^2 \tag{5.4.4}$$

Once again, the resemblance with the electric field version is clear, with the only difference being that the constant appears in the denominator here, while its electric counterpart appears in the numerator.

Enhancing Inductors

When we discussed capacitors, we found that we could alter their energy-storing capabilities by putting a dielectric between their plates. We have a similar option for inductors. We previously discussed the concepts related to magnetic fields in various substances, learning that substances can react in basically one of two ways: The magnetic dipoles in the substance can align with the field, or new dipoles can be induced which (according to Lenz's law) align opposite to the field. The former we called paramagnetism (or, if the dipoles remain aligned after removing the field, ferromagnetism) and it augments the applied field. The latter we called diamagnetism, and it reduces the applied field. This contrasts with the case of electricity, where insulating materials can only react as dielectrics, and only act to reduce the field.

As in the case of electricity, we will introduce a physical constant known as the *permeability*, which, like the permittivity in the case of electricity, takes the place of the vacuum constant:

$$\text{electricity: } \epsilon_o \rightarrow \epsilon \qquad \text{magnetism: } \mu_o \rightarrow \mu \tag{5.4.5}$$

As with the case of electricity, we simply replace the vacuum permeability with the permeability of the substance to get the answers inside of matter. So Biot-Savart and Ampére's laws are easily translated into more general forms:

$$\overrightarrow{B} = \frac{\mu}{4\pi}\int \frac{I\overrightarrow{dl}\times\hat{r}}{r^2} \qquad \oint \overrightarrow{B}\cdot\overrightarrow{dl} = \mu I_{enclosed} \tag{5.4.6}$$

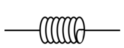
Notice from Biot-Savart's law that increasing the permeability for the same source *increases* the field strength (contrast this with the permittivity in the case of the Coulomb field). Therefore a permeability higher than the vacuum means the material is paramagnetic (and much higher than that is ferromagnetic). A value lower than the vacuum value corresponds to diamagnetism. Frequently materials are classified according to the percentage increase/decrease they provide to the total field compared to the vacuum case. That is:

$$\mu = (1+\chi_m)\,\mu_o \tag{5.4.7}$$

The constant χ_m is called the *magnetic susceptibility*. This has a positive for substances that are paramagnetic and ferromagnetic, and a negative value for diamagnetic substances.

LR Circuits

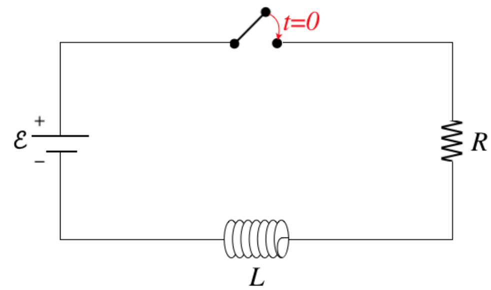
It's time to add inductors into our circuit diagrams, so we need a new symbol:

inductor: 

As with any other object in a circuit, there will be a specific voltage drop across the device as we invoke Kirchhoff's loop rule. The difference with this device it that its "smart battery" property makes it somewhat trickier than the other objects to determine the sign of the voltage change.

One reason to include an inductor in a circuit is to protect the circuit from current spikes (i.e. as a *surge protector*). If the current changes dramatically and suddenly, then the inductor will respond by providing an emf that opposes the sudden change, reducing the amount that the current is able to change over a short period, protecting the system from potential damage. We will see some other effects that an inductor has on a circuit as well, starting with how it interacts in a circuit with a resistor.

Figure 5.4.2a – An LR Circuit with Growing Current



When the switch is first closed, the current "wants" to jump instantly from zero to satisfy $\mathcal{E} = IR$, but the inductor doesn't allow this, because it develops an emf to oppose sudden changes. We begin with the Kirchhoff loop rule (which provides a new challenge for us when it comes to inductors), then solve the differential equation as we did for the RC circuit previously. To use the loop rule, we need to label a current and choose a loop direction. For the case above, let's choose clockwise for both. Going around this loop, the battery provides a voltage increase of $+\mathcal{E}$, and the resistor a voltage drop of $-IR$. What about the inductor?

When the switch is closed, the current that points right-to-left for the inductor increases in the direction of the loop. As a result of Faraday's law, the inductor becomes a "smart battery" that acts to reduce the current, which means there is a voltage *drop*:

$$\mathcal{E}_{inductor} = -L\frac{dI}{dt} \tag{5.4.8}$$

With the current increasing, the derivative is positive, and since L is always positive, a voltage drop requires a minus sign. Before we put the loop equation together, let's ask how this might change if we had labeled the current differently or chosen a different loop direction. First, if we switch the direction of the current label to left-to-right, and leave the loop direction, then an *increasing* current will result in the left side of the "smart battery" being at higher potential, which means that in a clockwise loop, the inductor would give a potential increase, and we would have to use $\mathcal{E}_{inductor} = +L\frac{dI}{dt}$. So it seems clear that we get the correct sign when we use the same convention as with the resistor – a minus sign when the current direction matches that of the loop direction, and a positive sign when the loop and current directions are opposite to each other.

Okay, so let's put together our loop equation and solve:

$$+\mathcal{E} - IR - L\frac{dI}{dt} = 0 \tag{5.4.9}$$

We have obtained a solution to this differential equation before (with different variables) – Equation 3.5.8. Following the same procedure to integrate this equation gives the result:

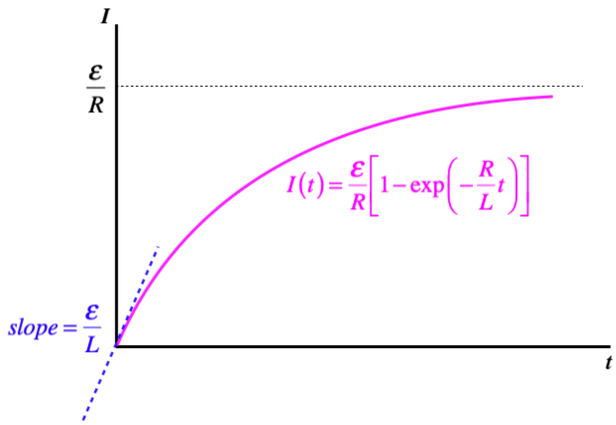
$$I(t) = \frac{\mathcal{E}}{R}\left(1 - e^{-\frac{t}{\tau}}\right) \;, \qquad \tau \equiv \frac{L}{R} \tag{5.4.10}$$

Note that the time constant for this circuit is quite different from the one for the RC circuit. Most notably, higher resistance in an RC circuit results in a larger time constant – it takes longer for the charge to decay from the plates of the capacitor when the resistance is higher, because it keeps the rate of flow (current) lower. In this case, however, a larger resistance causes the current to decay faster (i.e. $\frac{dI}{dt}$ is a more negative number):

$$\frac{dI}{dt} = \frac{1}{L}(\mathcal{E} - IR) \tag{5.4.11}$$

Faster decay means a smaller time constant.

Figure 5.4.2b – An LR Circuit with Growing Current



Let's check the extreme ends of this curve, to see if it makes sense. When the switch is first closed, the current grows at its greatest rate, *but this is not infinity*. That is, the current doesn't immediately jump to the value given by Ohm's law. The greater the inductance, the slower the initial growth in current is, since the slope of the current curve at $t=0$ is inversely-proportional to L. After a long time, the current-vs.-time curve flattens-out, and when the slope is zero, there is no emf induced in the inductor, which means that the current reaches the Ohm's law value – it gets to this point asymptotically.

Note that we can also witness this process in reverse – a circuit with an established current from which the battery is suddenly removed. In this case, we simply remove the $\mathcal{E}$ term from the differential equation, and the result is exponential decay, like a discharging capacitor. The time constant for this case is the same as the case of growing current:

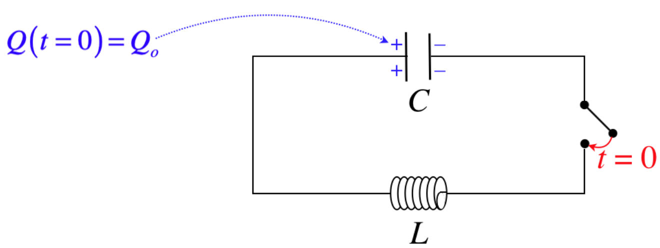
$$I(t)=I_oe^{-\frac{t}{\tau}}\;,\qquad \tau\equiv\frac{L}{R} \tag{5.4.12}$$

In terms of energy, it is easy to see what is going on here. The energy stored in the magnetic field is gradually converted into thermal energy energy by the resistor.

LC Circuits

Let's see what happens when we pair an inductor with a capacitor.

Figure 5.4.3 – An LC Circuit



Choosing the direction of the current through the inductor to be left-to-right, and the loop direction counterclockwise, we have:

$$+\frac{Q}{C}-L\frac{dI}{dt} \tag{5.4.13}$$

Next we have to recall how to relate the charge on the capacitor to the current. When this current is positive, charge is *leaving* the capacitor, which means that a decrease in Q is related to a positive value of I according to:

$$I=-\frac{dQ}{dt} \tag{5.4.14}$$

Putting this in above gives the differential equation:

$$\frac{d^2Q}{dt^2}+\frac{1}{LC}Q=0 \tag{5.4.15}$$

This is another differential equation we have seen before, though it was not in this class. Yes, this is the same differential equation that comes about for a mass oscillating on a spring. The solution for $Q(t)$ needs to be sinusoidal, since two derivatives of a sine or cosine function gives back a negative of itself (multiplied by a constant that comes from the chain rule). The solution to this particular case (with the starting charge at $t = 0$ given) is:

$$Q(t) = Q_o \cos(\omega t)\ , \qquad \omega \equiv \frac{1}{\sqrt{LC}} \tag{5.4.16}$$

Interpreting this result, we see that the charge actually sloshes back-and-forth between the plates (the charges on the plates actually eventually swap places!). We can also write down the equation for the current:

$$I(t) = -\frac{dQ}{dt} = Q_o\omega\sin(\omega t) = \frac{Q_o}{\sqrt{LC}}\sin(\omega t) = I_{max}\sin(\omega t)\ , \qquad I_{max} \equiv \frac{Q_o}{\sqrt{LC}} \tag{5.4.17}$$

We see that the current starts at zero, and grows to a maximum value, and this maximum occurs when the value of the sine is 1, which is the same time that the charge on the capacitor reaches zero. This actually gives us insight into the energy considerations for this circuit. Energy isn't being converted to thermal energy by a resistor, so it has no way to exit, which means that the oscillations continue indefinitely. We know exactly how much energy the circuit starts with:

$$U_{tot} = \frac{Q_o^2}{2C} \tag{5.4.18}$$

When all of the charge is gone, the current hits a maximum, which means that all of the energy is then in the magnetic field. It's easy to confirm that the energy is conserved:

$$U_{tot} = \frac{1}{2}LI_{max}^2 = \frac{1}{2}L\left(\frac{Q_o}{\sqrt{LC}}\right)^2 = \frac{Q_o^2}{2C} \tag{5.4.19}$$

Example 5.4.1

Show that the total energy in the LC circuit remains unchanged at all times, not just when all the energy is in the capacitor or inductor.

Solution

The energy stored in the system at a time t is the sum of the energies stored in each device:

$$U(t) = \frac{1}{2C}[Q(t)]^2 + \frac{1}{2}L[I(t)]^2 = \frac{1}{2C}[Q_o\cos(\omega t)]^2 + \frac{1}{2}L[I_{max}\sin(\omega t)]^2$$

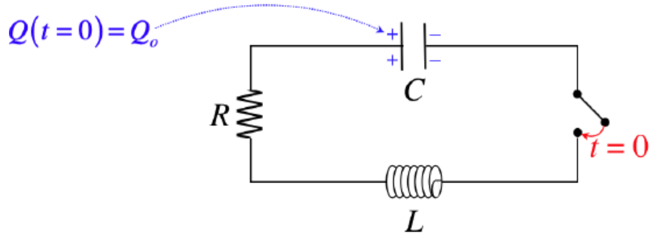
We have already established that the maximum values are equal, so:

$$\frac{1}{2}LI_{max}^2 = \frac{Q_0^2}{2C} \quad\Rightarrow\quad U(t) = \frac{Q_0^2}{2C}\left[\cos^2(\omega t) + \sin^2(\omega t)\right] = \frac{Q_0^2}{2C}$$

LRC Circuits

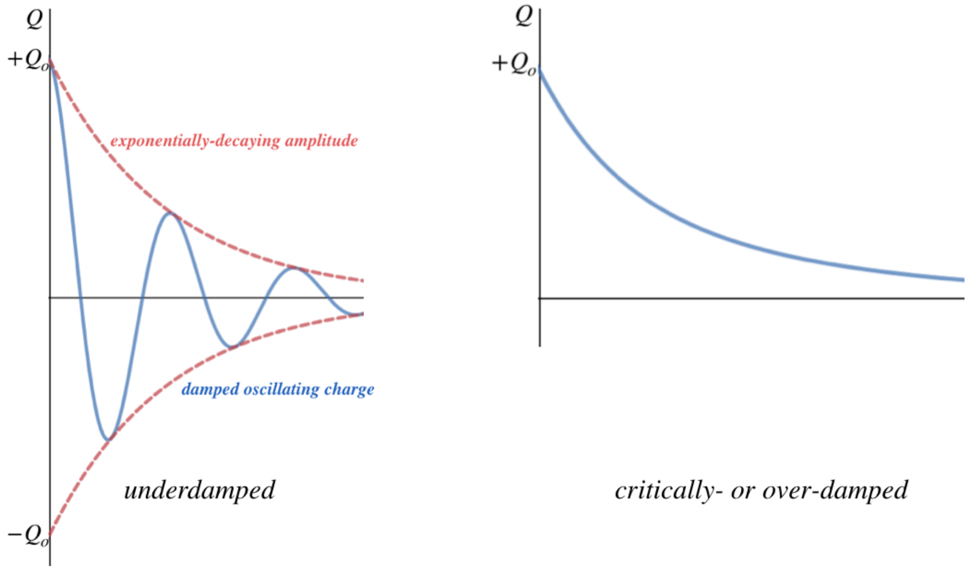
All that remains to examine in terms of circuits that combine different components is to put all of them together. We can guess the result – the resistance results in decay, as the energy in the circuit gets converted to thermal. The capacitance and inductance do their dance of oscillation between electric and magnetic field energy. Putting them all together results in the equivalent of a damped oscillator (a harmonic oscillator with friction).

Figure 5.4.4 – An LRC Circuit



To get to this result, we (as usual) start with Kirchhoff's loop rule. This time the solution to differential equation has different characteristics, depending upon the values of the constants involved. For example, if the resistance is above a certain amount, the current dissipates before the charge is able to switch plates on the capacitor – it just decays directly down to zero. This is called an *overdamped* system. If the resistance is *just barely* large enough to cause this behavior, the system is said to be *critically-damped*. And if the resistance is low enough to allow oscillation, it is called *under-damped*. In this case, the charge does oscillate between the two capacitor plates, filling them a little less with every iteration.

Figure 5.4.5 – Current Behavior Based on Circuit Details



The critical criterion for determining which of these occurs is a comparison of R^2 and $\frac{4L}{C}$:

$$\begin{aligned} \text{underdamped:} \quad & R^2 < \frac{4L}{C} \\ \text{critically-damped:} \quad & R^2 = \frac{4L}{C} \\ \text{overdamped:} \quad & R^2 > \frac{4L}{C} \end{aligned} \tag{5.4.20}$$

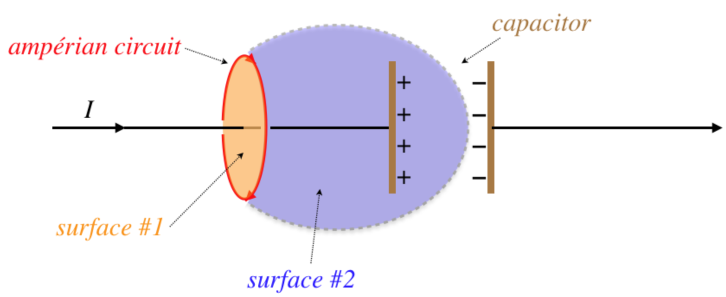
5.5: Maxwell's Equations

Ampére's Law is Broken

As far as the EM theory had come, a 19th-century Scottish physicist named James Clerk Maxwell felt something had to be missing. To get an idea of what was nagging him, consider Ampére's law. Recall we said that it only worked for closed loops and infinitely-long wires, because the current had to pierce a surface bounded by the Ampérian circuit. Maxwell felt that there had to be some way to modify Ampére's law to take care of this shortcoming, and came up with the following thought experiment.

Suppose we have a long-straight wire with a current in it. We can employ Ampére's law for this situation, because if we construct an Ampérian circuit around this wire, every surface – whatever its shape – that is bounded by that closed path must be pierced by that wire. Now suppose the wire includes a single capacitor. Now it is possible to construct a surface bounded by the Ampérian circuit such that the current does *not* pierce it. Maxwell felt that maybe maybe Ampére's law could be modified such that surfaces not pierced by the current can also be related to the line integral of the magnetic field around the same circuit.

Figure 5.5.1 – Maxwell's Extension of Ampére's law



The current piercing surface #1 can be expressed in a manner that transports (or "displaces") the calculation over to the capacitor's electric field. The current passing through surface #1 is the rate at which the charge is building up on the capacitor plate, and this is reated to the rate at which the field between the two capacitor plates is growing. Specifically:

$$\left.\begin{array}{l} I=\dfrac{dQ}{dt} \\ \left|\overrightarrow{E}\right|=\dfrac{\sigma}{\epsilon_o}=\dfrac{Q}{\epsilon_o A}\end{array}\right\} \;\; I=\frac{d}{dt}\left(\epsilon_o\left|\overrightarrow{E}\right|A\right)=\epsilon_o\frac{d}{dt}\int\limits_{\text{surface }\#2}\overrightarrow{E}\cdot d\overrightarrow{A} \tag{5.5.1}$$

The time rate of change of the electric field flux which accounts for the enclosed current for a surface that is displaced was called the *displacement current* by Maxwell. It accounts for the fact that while charge does not pass through a particular surface over time, an equivalent about of "current" in the form of increasing (or decreasing) electric field flux takes its place. So in general, in cases where there is both a current piercing a surface *and* a change in the electric flux through that surface, the line integral of the magnetic field around a closed path that borders that surface is:

$$\oint\overrightarrow{B}\cdot\overrightarrow{dl}=\mu_o\left(I_{moving\;charge}+Idisplacement\right)=\mu_o I_{encl}+\mu_o\epsilon_o\frac{d}{dt}\int\overrightarrow{E}\cdot d\overrightarrow{A} \tag{5.5.2}$$

This is Ampére's law modified with Maxwell's displacement current in integral form. We found earlier that Ampére's law could be written in local (differential) form using Stoke's theorem, and since the integral of the electric field flux is over a surface bounded by the same closed path, we can include the second term in this equation:

$$\oint\overrightarrow{B}\cdot\overrightarrow{dl}=\int\left(\overrightarrow{\nabla}\times\overrightarrow{B}\right)\cdot d\overrightarrow{A}=\mu_o\int\overrightarrow{J}\cdot d\overrightarrow{A}+\mu_o\epsilon_o\frac{d}{dt}\int\overrightarrow{E}\cdot d\overrightarrow{A}\quad\Rightarrow\quad\overrightarrow{\nabla}\times\overrightarrow{B}=\mu_o\overrightarrow{J}+\mu_o\epsilon_o\frac{d}{dt}\overrightarrow{E} \tag{5.5.3}$$

Notice that in fact we will get a nonzero magnetic field line integral *even if there is no moving charge*, if there is a time-varying electric field present. What Maxwell had discovered was, not only did Faraday's law tell us that a time-varying magnetic field causes an electric field to circulate around it, but it worked in the other direction as well: a time-varying electric field gives rise to a magnetic field as well.

Summary of Field Equations

We can now put *all* of the field equations together, in both integral and local form, to construct a complete theory of electromagnetism. It is summarized in four equations, now known as *Maxwell's equations*:

Figure 5.5.2 – Maxwell's Equations

Physical Principle	*Integral Form*	*Vector Calculus Link*	*Differential (Local) Form*
Gauss's Law: inverse-square field of monopole	$\oint \vec{E}\cdot d\vec{A} = \frac{q_{encl}}{\varepsilon_o}$	Divergence Theorem $+\int \rho\, dV = q_{encl}$	$\vec{\nabla}\cdot\vec{E} = \frac{\rho}{\varepsilon_o}$
Gauss's Law: magnetic monopoles do not exist	$\oint \vec{B}\cdot d\vec{A} = 0$	Divergence Theorem	$\vec{\nabla}\cdot\vec{B} = 0$
Faraday's Law: changing magnetic field induces electric field	$\oint \vec{E}\cdot d\vec{l} = -\frac{d}{dt}\int \vec{B}\cdot d\vec{A}$	Stokes's Theorem	$\vec{\nabla}\times\vec{E} = -\frac{d\vec{B}}{dt}$
Ampére & Maxwell: magnetic field caused by current or changing E-field	$\oint \vec{B}\cdot d\vec{l} = \mu_o I_{encl} + \mu_o\varepsilon_o \frac{d}{dt}\int \vec{E}\cdot d\vec{A}$	Stokes's Theorem $+\int \vec{J}\cdot d\vec{A} = I_{encl}$	$\vec{\nabla}\times\vec{B} = \mu_o\vec{J} + \mu_o\varepsilon_o\frac{d\vec{E}}{dt}$

Charge Conservation

Electric charge conservation is a fundamental element of the theory of electromagnetism, which we first addressed at the end of Section 3.1, culminating in Equation 3.1.8. Electric charges as sources of both fields are included in Maxwell's equations, so it is absolutely essential that Maxwell's equations be consistent with charge conservation. Thanks to MAxwell's contribution, charge conservation can be *derived* from the field equations. To see this, consider the identity we have mentioned previously – that the divergence of the curl of any vector field vanishes. Applying this identity to the Ampére/Maxwell equation gives:

$$0 = \overrightarrow{\nabla}\cdot\left(\overrightarrow{\nabla}\times\overrightarrow{B}\right) = \mu_o\overrightarrow{\nabla}\cdot\overrightarrow{J} + \mu_o\epsilon_o\overrightarrow{\nabla}\cdot\frac{d\overrightarrow{E}}{dt} = \overrightarrow{\nabla}\cdot\overrightarrow{J} + \frac{d}{dt}\left(\epsilon_o\overrightarrow{\nabla}\cdot\overrightarrow{E}\right) \tag{5.5.4}$$

Now applying the local form of Gauss's law for electric fields to the last term gives the continuity equation (Equation 3.1.8), which expresses charge conservation:

$$0 = \overrightarrow{\nabla}\cdot\overrightarrow{J} + \frac{d\rho}{dt} \tag{5.5.5}$$

So essentially charge conservation is "baked into" the field equations. The field equations give a complete accounting of how fields are generated from conserved electric charge (and its motion), and how the two types of field (electric and magnetic) are generated from each other. What they do not provide is how electric charge is *affected* by the fields, so we need to add–in the Lorentz force (Equation 4.1.6) to complete the theory:

$$\overrightarrow{F} = q\left(\overrightarrow{E} + \overrightarrow{v}\times\overrightarrow{B}\right) \tag{5.5.6}$$

It turns out that rather than provide the Lorentz force, the interactions of charges with fields can be obtained by knowing the energy densities of the fields, in a manner similar to deriving force from the gradient of potential energy. That is, the theory is also complete if instead of the Lorentz force, one knows:

$$U_{EM} = \frac{1}{2}\epsilon_o E^2 + \frac{1}{2\mu_o}B^2 \tag{5.5.7}$$

5.6: Electromagnetic Waves

The Wave Equation

When Maxwell realized that his new addition to the theory meant that not only can changing magnetic fields induce electric fields (Faraday), but changing electric fields can also induce magnetic fields, it occurred to him that it might be possible for propagation to occur: A changing magnetic field creates a changing electric field, which creates a changing magnetic field, and so on.

It was not hard for a mathematician such as Maxwell to express this propagation mathematically. To see how it comes about, let's simplify our physical situation by considering a region free of charges. This results in a simplified set of Maxwell's equations:

$$\begin{array}{ll} \text{electric Gauss:} & \overrightarrow{\nabla}\cdot\overrightarrow{E}=0 \\ \text{magnetic Gauss:} & \overrightarrow{\nabla}\cdot\overrightarrow{B}=0 \\ \text{Faraday:} & \overrightarrow{\nabla}\times\overrightarrow{E}=-\dfrac{d}{dt}\overrightarrow{B} \\ \text{Maxwell:} & \overrightarrow{\nabla}\times\overrightarrow{B}=\mu_o\epsilon_o\dfrac{d}{dt}\overrightarrow{E} \end{array} \tag{5.6.1}$$

Let's start by taking a derivative of the equation of the Maxwell equation with respect to time:

$$\frac{d}{dt}\overrightarrow{\nabla}\times\overrightarrow{B}=\overrightarrow{\nabla}\times\frac{d}{dt}\overrightarrow{B}=\mu_o\epsilon_o\frac{d^2}{dt^2}\overrightarrow{E} \tag{5.6.2}$$

Now plug the equation of Faraday into the derivative of the magnetic field:

$$\overrightarrow{\nabla}\times\left(-\overrightarrow{\nabla}\times\overrightarrow{E}\right)=\mu_o\epsilon_o\frac{d^2}{dt^2}\overrightarrow{E} \tag{5.6.3}$$

Now we have an equation exclusively in terms of the electric field (electric field induces magnetic field which induces electric field again). The double curl looks quite daunting to simplify, but it turns out that there is a useful identity from vector calculus to save the day:

$$\overrightarrow{\nabla}\times\left(\overrightarrow{\nabla}\times\overrightarrow{E}\right)=\overrightarrow{\nabla}\left(\overrightarrow{\nabla}\cdot\overrightarrow{E}\right)+\nabla^2\overrightarrow{E} \tag{5.6.4}$$

Plugging the electric Gauss equation into this and then plugging this equation in for the double curl gives:

$$\nabla^2\overrightarrow{E}=\mu_o\epsilon_o\frac{d^2\overrightarrow{E}}{dt^2} \tag{5.6.5}$$

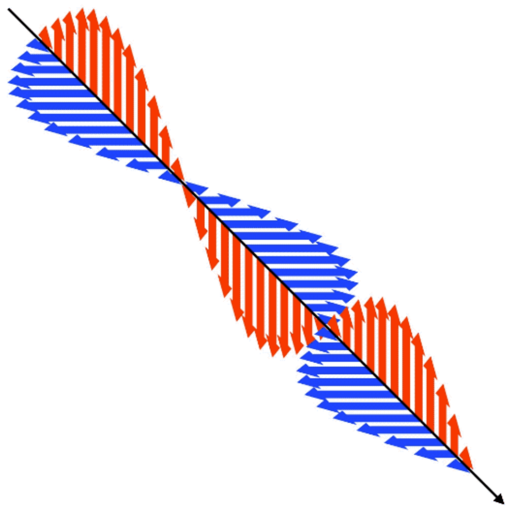
Perhaps you recognize this differential equation from Physics 9B? It is the wave equation – not surprising, really, given that a changing electric field seems to propagate another electric field (using the changing magnetic field as an intermediate step). Naturally Maxwell recognized the wave equation as well, and asked the most obvious question, "How fast is this wave?" Given that the velocity of a wave can be taken directly from the wave equation, this is not hard to calculate. The coefficient of the second time derivative term is the inverse of the square of the wave speed, so the speed of this wave is:

$$v=\frac{1}{\sqrt{\mu_o\epsilon_o}}=\frac{1}{\sqrt{\left(4\pi\times10^{-7}\frac{Ns^2}{C^2}\right)\left(8.85\times10^{-12}\frac{C^2}{Nm^2}\right)}}=3.0\times10^8\frac{m}{s} \tag{5.6.6}$$

Well of course Maxwell recognized this number immediately (as should you!) – it is the speed of light, c. Maxwell has shown that light is an electromagnetic phenomenon that exists because electric and magnetic fields can propagate by inducing each other.

If one begins the derivation above by taking a derivative of the Faraday equation with respect to time and follows the same steps, one finds that the very same wave equation applies to the magnetic field – both fields propagate together as a single light ("electromagnetic") wave.

Figure 5.6.1 – Electromagnetic Wave



EM Wave Properties

Let’s see what we can find out about these waves by looking at a specific example. Suppose we have a harmonic plane wave of electric field polarized in the x-z plane. Recall from 9B that this is expressed mathematically by:

$$\overrightarrow{E}(z,t)=\hat{i}\;E_o\cos\left(\frac{2\pi}{\lambda}z-\frac{2\pi}{T}t\right) \tag{5.6.7}$$

This represents a wave that propagates along the z direction, the "displacement" direction (polarization direction of the electric field vectors) along the x direction, has an amplitude of E_o, a wavelength of λ, and period of T. We have chosen the starting time such that the phase constant is zero.

Let's plug this field into Faraday's equation by taking its curl:

$$\overrightarrow{\nabla}\times\overrightarrow{E}=\begin{vmatrix}\hat{i} & \hat{j} & \hat{k}\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ E_x & E_y & E_z\end{vmatrix}=\begin{vmatrix}\hat{i} & \hat{j} & \hat{k}\\ 0 & 0 & \frac{\partial}{\partial z}\\ E & 0 & 0\end{vmatrix}=\frac{\partial E}{\partial z}\left(+\hat{j}\right) \tag{5.6.8}$$

Performing the derivative, we get:

$$\overrightarrow{\nabla}\times\overrightarrow{E}=-\left(\frac{2\pi}{\lambda}\right)E_o\sin\left(\frac{2\pi}{\lambda}z-\frac{2\pi}{T}t\right)\hat{j} \tag{5.6.9}$$

Now that we know the curl of the electric field, we can plug the result into Faraday's law:

$$\overrightarrow{\nabla}\times\overrightarrow{E}=-\frac{d}{dt}\overrightarrow{B}=-\left(\frac{2\pi}{\lambda}\right)E_o\sin\left(\frac{2\pi}{\lambda}z-\frac{2\pi}{T}t\right)\hat{j} \tag{5.6.10}$$

We can now integrate to find the wave function for the magnetic field of this wave (for simplicity, we will assume that the electric and magnetic fields are in phase with each other, which will mean the arbitrary constant from the integral is just zero):

$$\vec{B}(z,t) = +\hat{j}\left(\frac{2\pi}{\lambda}\right)E_o\int\sin\left(\frac{2\pi}{\lambda}z - \frac{2\pi}{T}t\right)dt = +\hat{j}\left(\frac{T}{\lambda}\right)E_o\cos\left(\frac{2\pi}{\lambda}z - \frac{2\pi}{T}t\right) \tag{5.6.11}$$

We see that the magnetic field wave function has the same frequency and wavelength as the electric field wave function, and since the ratio $\frac{T}{\lambda}$ is just the inverse of the speed of the wave c, which means that the amplitudes of the electric and magnetic parts of the wave are related by:

$$B_o = \frac{E_o}{c} \tag{5.6.12}$$

We can also see a how the various directions are related. The velocity is in the $\hat{k}$ direction, the electric field in the $\hat{i}$ direction, and the magnetic field in the $\hat{j}$ direction – all three of these vectors are mutually orthogonal. In fact, the direction of the wave's velocity vector is the same direction as the vector $\vec{E}\times\vec{B}$.

Example 5.6.1

We know that electric and magnetic fields store energy in the space in which they exist. As a light wave passes through a region of space, the fluctuating fields cause the energy density in that space to fluctuate. Is more of the wave's energy a result of the electric field or the magnetic field? More specifically, compute the ratio of the maximum energy densities of the two fields within a single EM wave traveling through a vacuum.

Solution

The energy densities for electric and magnetic fields in a vacuum are given by:

$$U_E = \tfrac{1}{2}\epsilon_o E^2$$
$$U_B = \tfrac{1}{2\mu_o}B^2$$

The maximum energy densities come about when the fields equal their amplitudes, so taking the ratio of these energies gives:

$$\frac{U_E}{U_B} = \frac{\epsilon_o E^2}{\frac{1}{\mu_o}B^2} = \epsilon_o\mu_o\frac{E_o^2}{B_o^2}$$

Now plugging in Equation 5.6.6 and Equation 5.6.12, we get the simple result:

$$U_E = U_B$$

Both fields contribute equally to the energy density in the space through which the wave passes.

Glossary

Sample Word 1 | Sample Definition 1