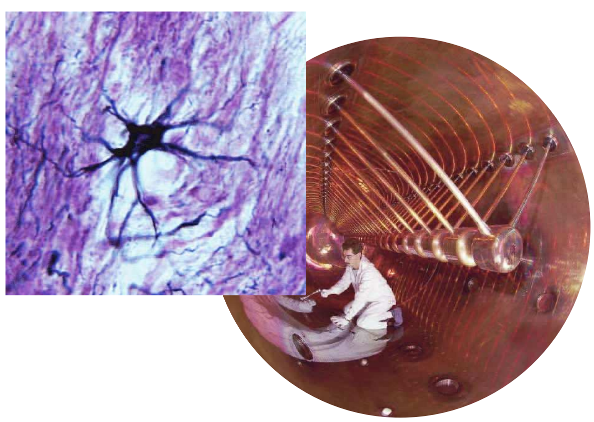
# Electricity and Magnetism

## Benjamin Crowell

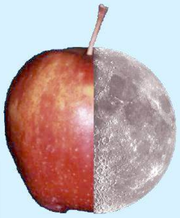


Book 4 in the Light and Matter series of free introductory physics textbooks
www.lightandmatter.com

# Electricity and Magnetism

The **Light and Matter** series of introductory physics textbooks:

1 Newtonian Physics
2 Conservation Laws
3 Vibrations and Waves
4 Electricity and Magnetism
5 Optics
6 The Modern Revolution in Physics



# Electricity and Magnetism

Benjamin Crowell

www.lightandmatter.com

Light and Matter
Fullerton, California
www.lightandmatter.com

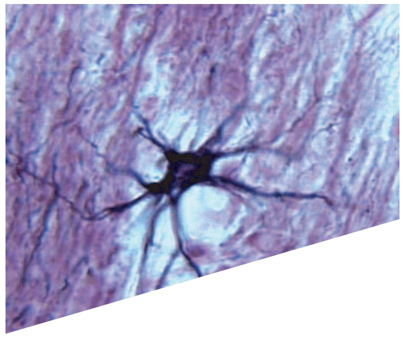


edition 2.3
rev. 21st October 2006



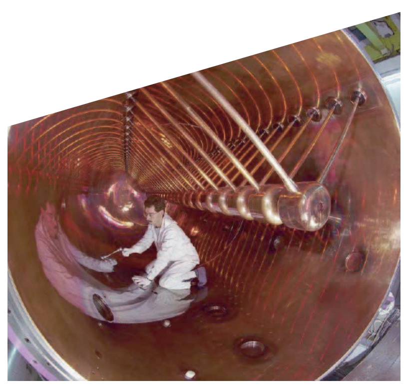
ISBN 0-9704670-4-4

To Arnold Arons.



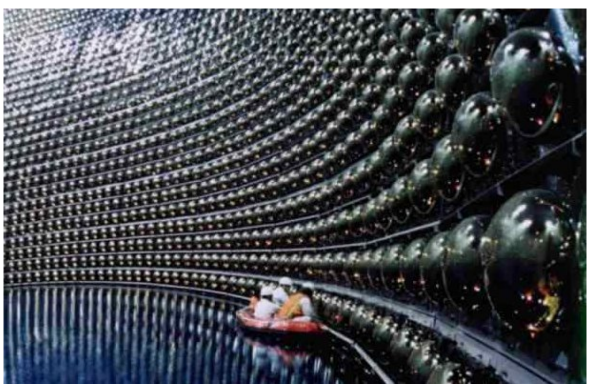
# Brief Contents



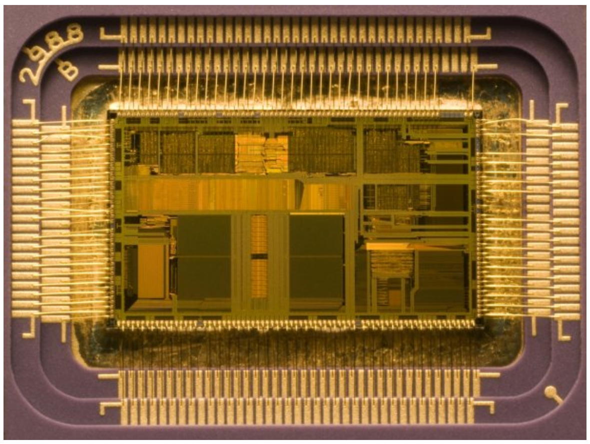


# Contents











HUYGENS

# Chapter 1

# Electricity and the Atom

Where the telescope ends, the microscope begins. Which of the two has the grander view? *Victor Hugo*

His father died during his mother's pregnancy. Rejected by her as a boy, he was packed off to boarding school when she remarried. He himself never married, but in middle age he formed an intense relationship with a much younger man, a relationship that he terminated when he underwent a psychotic break. Following his early scientific successes, he spent the rest of his professional life mostly in frustration over his inability to unlock the secrets of alchemy.

The man being described is Isaac Newton, but not the triumphant Newton of the standard textbook hagiography. Why dwell on the sad side of his life? To the modern science educator, Newton's lifelong obsession with alchemy may seem an embarrassment, a distraction from his main achievement, the creation the modern science of mechanics. To Newton, however, his alchemical researches were naturally related to his investigations of force and motion. What was radical about Newton's analysis of motion was its universality: it succeeded in describing both the heavens and the earth with the same equations, whereas previously it had been assumed that the sun, moon, stars, and planets were fundamentally different from earthly objects. But Newton realized that if science was to describe all of nature in a unified way, it was not enough to unite the human scale with the scale of the universe: he would not be satisfied until he fit the microscopic universe into the picture as well.

It should not surprise us that Newton failed. Although he was a firm believer in the existence of atoms, there was no more experimental evidence for their existence than there had been when the ancient Greeks first posited them on purely philosophical grounds. Alchemy labored under a tradition of secrecy and mysticism. Newton had already almost single-handedly transformed the fuzzyheaded field of "natural philosophy" into something we would recognize as the modern science of physics, and it would be unjust to criticize him for failing to change alchemy into modern chemistry as well. The time was not ripe. The microscope was a new invention, and it was cutting-edge science when Newton's contemporary Hooke discovered that living things were made out of cells.

## 1.1 The quest for the atomic force

Newton was not the first of the age of reason. He was the last of the magicians. *John Maynard Keynes*

Nevertheless it will be instructive to pick up Newton's train of thought and see where it leads us with the benefit of modern hindsight. In uniting the human and cosmic scales of existence, he had reimagined both as stages on which the actors were objects (trees and houses, planets and stars) that interacted through attractions and repulsions. He was already convinced that the objects inhabiting the microworld were atoms, so it remained only to determine what kinds of forces they exerted on each other.

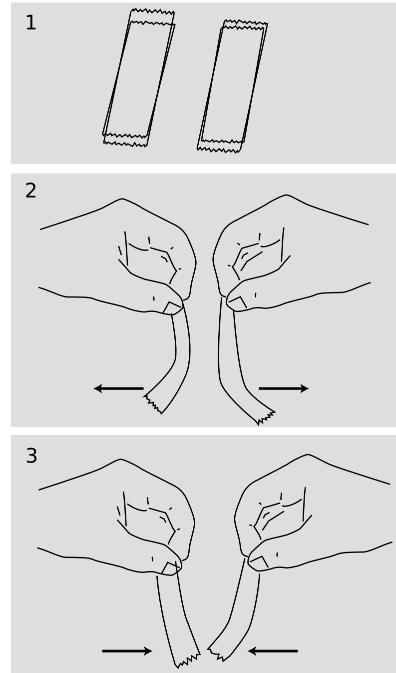
His next insight was no less brilliant for his inability to bring it to fruition. He realized that the many human-scale forces — friction, sticky forces, the normal forces that keep objects from occupying the same space, and so on — must all simply be expressions of a more fundamental force acting between atoms. Tape sticks to paper because the atoms in the tape attract the atoms in the paper. My house doesn't fall to the center of the earth because its atoms repel the atoms of the dirt under it.

Here he got stuck. It was tempting to think that the atomic force was a form of gravity, which he knew to be universal, fundamental, and mathematically simple. Gravity, however, is always attractive, so how could he use it to explain the existence of both attractive and repulsive atomic forces? The gravitational force between objects of ordinary size is also extremely small, which is why we never notice cars and houses attracting us gravitationally. It would be hard to understand how gravity could be responsible for anything as vigorous as the beating of a heart or the explosion of gunpowder. Newton went on to write a million words of alchemical notes filled with speculation about some other force, perhaps a "divine force" or "vegetative force" that would for example be carried by the sperm to the egg.

Luckily, we now know enough to investigate a different suspect as a candidate for the atomic force: electricity. Electric forces are often observed between objects that have been prepared by rubbing (or other surface interactions), for instance when clothes rub against each other in the dryer. A useful example is shown in figure a/1: stick two pieces of tape on a tabletop, and then put two more pieces on top of them. Lift each pair from the table, and then separate them. The two top pieces will then repel each other, a/2, as will the two bottom pieces. A bottom piece will attract a top piece, however, a/3. Electrical forces like these are similar in certain ways to gravity, the other force that we already know to be fundamental:

- Electrical forces are *universal*. Although some substances, such as fur, rubber, and plastic, respond more strongly to electrical preparation than others, all matter participates in electrical forces to some degree. There is no such thing as a "nonelectric" substance. Matter is both inherently gravitational and inherently electrical.
- Experiments show that the electrical force, like the gravitational force, is an *inverse square* force. That is, the electrical force between two spheres is proportional to $1/r^2$, where $r$ is the center-to-center distance between them.

Furthermore, electrical forces make more sense than gravity as candidates for the fundamental force between atoms, because we have observed that they can be either attractive or repulsive.




a / Four pieces of tape are prepared, 1, as described in the text. Depending on which combination is tested, the interaction can be either repulsive, 2, or attractive, 3.

## 1.2 Charge, electricity and magnetism

### Charge

"Charge" is the technical term used to indicate that an object has been prepared so as to participate in electrical forces. This is to be distinguished from the common usage, in which the term is used indiscriminately for anything electrical. For example, although we speak colloquially of "charging" a battery, you may easily verify that a battery has no charge in the technical sense, e.g., it does not exert any electrical force on a piece of tape that has been prepared as described in the previous section.

*Two types of charge*

We can easily collect reams of data on electrical forces between different substances that have been charged in different ways. We find for example that cat fur prepared by rubbing against rabbit fur will attract glass that has been rubbed on silk. How can we make any sense of all this information? A vast simplification is achieved by noting that there are really only two types of charge.

Suppose we pick cat fur rubbed on rabbit fur as a representative of type A, and glass rubbed on silk for type B. We will now find that there is no "type C." Any object electrified by any method is either A-like, attracting things A attracts and repelling those it repels, or B-like, displaying the same attractions and repulsions as B. The two types, A and B, always display opposite interactions. If A displays an attraction with some charged object, then B is guaranteed to undergo repulsion with it, and vice-versa.

*The coulomb*

Although there are only two types of charge, each type can come in different amounts. The metric unit of charge is the coulomb (rhymes with "drool on"), defined as follows:

> One Coulomb (C) is the amount of charge such that a force of $9.0 \times 10^9$ N occurs between two pointlike objects with charges of 1 C separated by a distance of 1 m.

The notation for an amount of charge is $q$. The numerical factor in the definition is historical in origin, and is not worth memorizing. The definition is stated for pointlike, i.e., very small, objects, because otherwise different parts of them would be at different distances from each other.

*A model of two types of charged particles*

Experiments show that all the methods of rubbing or otherwise charging objects involve two objects, and both of them end up getting charged. If one object acquires a certain amount of one type of charge, then the other ends up with an equal amount of the other type. Various interpretations of this are possible, but the simplest is that the basic building blocks of matter come in two flavors, one with each type of charge. Rubbing objects together results in the transfer of some of these particles from one object to the other. In this model, an object that has not been electrically prepared may actually possesses a great deal of *both* types of charge, but the amounts are equal and they are distributed in the same way throughout it. Since type A repels anything that type B attracts, and vice versa, the object will make a total force of zero on any other object. The rest of this chapter fleshes out this model and discusses how these mysterious particles can be understood as being internal parts of atoms.

*Use of positive and negative signs for charge*

Because the two types of charge tend to cancel out each other's forces, it makes sense to label them using positive and negative signs, and to discuss the *total* charge of an object. It is entirely arbitrary which type of charge to call negative and which to call positive. Benjamin Franklin decided to describe the one we've been calling "A" as negative, but it really doesn't matter as long as everyone is consistent with everyone else. An object with a total charge of zero (equal amounts of both types) is referred to as electrically *neutral*.

*self-check A*

Criticize the following statement: "There are two types of charge, attractive and repulsive." ▷ Answer, p. 195

*Coulomb's law*

A large body of experimental observations can be summarized as follows:

Coulomb's law: The magnitude of the force acting between pointlike charged objects at a center-to-center distance $r$ is given by the equation

$$|\mathbf{F}| = k\frac{|q_1||q_2|}{r^2} \qquad ,$$

where the constant $k$ equals $9.0\times10^9$ N·m$^2$/C$^2$. The force is attractive if the charges are of different signs, and repulsive if they have the same sign.

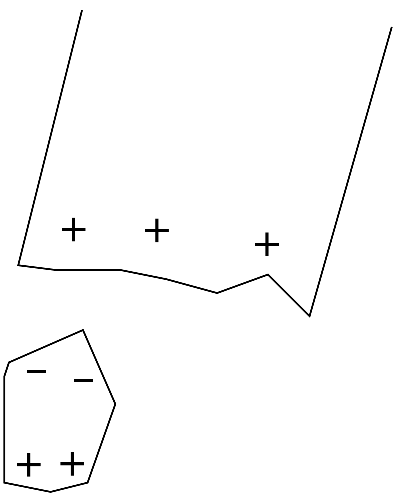
Clever modern techniques have allowed the $1/r^2$ form of Coulomb's law to be tested to incredible accuracy, showing that the exponent is in the range from 1.9999999999999998 to 2.0000000000000002.

Note that Coulomb's law is closely analogous to Newton's law of gravity, where the magnitude of the force is $Gm_1m_2/r^2$, except that there is only one type of mass, not two, and gravitational forces are never repulsive. Because of this close analogy between the two types of forces, we can recycle a great deal of our knowledge of gravitational forces. For instance, there is an electrical equivalent of the shell theorem: the electrical forces exerted externally by a uniformly charged spherical shell are the same as if all the charge was concentrated at its center, and the forces exerted internally are zero.

## Conservation of charge

An even more fundamental reason for using positive and negative signs for electrical charge is that experiments show that charge is conserved according to this definition: in any closed system, the total amount of charge is a constant. This is why we observe that rubbing initially uncharged substances together always has the result that one gains a certain amount of one type of charge, while the other acquires an equal amount of the other type. Conservation of charge seems natural in our model in which matter is made of positive and negative particles. If the charge on each particle is a fixed property of that type of particle, and if the particles themselves can be neither created nor destroyed, then conservation of charge is inevitable.

b / A charged piece of tape attracts uncharged pieces of paper from a distance, and they leap up to it.




c / The paper has zero total charge, but it does have charged particles in it that can move.

### Electrical forces involving neutral objects

As shown in figure b, an electrically charged object can attract objects that are uncharged. How is this possible? The key is that even though each piece of paper has a total charge of zero, it has at least some charged particles in it that have some freedom to move. Suppose that the tape is positively charged, c. Mobile particles in the paper will respond to the tape's forces, causing one end of the paper to become negatively charged and the other to become positive. The attraction is between the paper and the tape is now stronger than the repulsion, because the negatively charged end is closer to the tape.

*self-check B*

What would have happened if the tape was negatively charged? ▹ Answer, p. 195

### The path ahead

We have begun to encounter complex electrical behavior that we would never have realized was occurring just from the evidence of our eyes. Unlike the pulleys, blocks, and inclined planes of mechanics, the actors on the stage of electricity and magnetism are invisible phenomena alien to our everyday experience. For this reason, the flavor of the second half of your physics education is dramatically different, focusing much more on experiments and techniques. Even though you will never actually see charge moving through a wire, you can learn to use an ammeter to measure the flow.

Students also tend to get the impression from their first semester of physics that it is a dead science. Not so! We are about to pick up the historical trail that leads directly to the cutting-edge physics research you read about in the newspaper. The atom-smashing experiments that began around 1900, which we will be studying in this chapter, were not that different from the ones of the year 2000 — just smaller, simpler, and much cheaper.

### Magnetic forces

A detailed mathematical treatment of magnetism won't come until much later in this book, but we need to develop a few simple ideas about magnetism now because magnetic forces are used in the experiments and techniques we come to next. Everyday magnets come in two general types. Permanent magnets, such as the ones on your refrigerator, are made of iron or substances like steel that contain iron atoms. (Certain other substances also work, but iron is the cheapest and most common.) The other type of magnet, an example of which is the ones that make your stereo speakers vibrate, consist of coils of wire through which electric charge flows. Both types of magnets are able to attract iron that has not been magnetically prepared, for instance the door of the refrigerator.

A single insight makes these apparently complex phenomena much simpler to understand: magnetic forces are interactions between moving charges, occurring in addition to the electric forces. Suppose a permanent magnet is brought near a magnet of the coiled-wire type. The coiled wire has moving charges in it because we force charge to flow. The permanent magnet also has moving charges in it, but in this case the charges that naturally swirl around inside the iron. (What makes a magnetized piece of iron different from a block of wood is that the motion of the charge in the wood is random rather than organized.) The moving charges in the coiled-wire magnet exert a force on the moving charges in the permanent magnet, and vice-versa.

The mathematics of magnetism is significantly more complex than the Coulomb force law for electricity, which is why we will wait until chapter 6 before delving deeply into it. Two simple facts will suffice for now:

(1) If a charged particle is moving in a region of space near where other charged particles are also moving, their magnetic force on it is directly proportional to its velocity.

(2) The magnetic force on a moving charged particle is always perpendicular to the direction the particle is moving.

*A magnetic compass* *example 1*

The Earth is molten inside, and like a pot of boiling water, it roils and churns. To make a drastic oversimplification, electric charge can get carried along with the churning motion, so the Earth contains moving charge. The needle of a magnetic compass is itself a small permanent magnet. The moving charge inside the earth interacts magnetically with the moving charge inside the compass needle, causing the compass needle to twist around and point north.

*A television tube* *example 2*

A TV picture is painted by a stream of electrons coming from the back of the tube to the front. The beam scans across the whole surface of the tube like a reader scanning a page of a book. Magnetic forces are used to steer the beam. As the beam comes from the back of the tube to the front, up-down and left-right forces are needed for steering. But magnetic forces cannot be used to get the beam up to speed in the first place, since they can only push perpendicular to the electrons' direction of motion, not forward along it.

### Discussion Questions

**A** If the electrical attraction between two pointlike objects at a distance of 1 m is $9\times10^9$ N, why can't we infer that their charges are +1 and $-1$ C? What further observations would we need to do in order to prove this?

**B** An electrically charged piece of tape will be attracted to your hand. Does that allow us to tell whether the mobile charged particles in your hand are positive or negative, or both?

## 1.3 Atoms

I was brought up to look at the atom as a nice, hard fellow, red or grey in color according to taste. *Rutherford*

### Atomism

The Greeks have been kicked around a lot in the last couple of millennia: dominated by the Romans, bullied during the crusades by warlords going to and from the Holy Land, and occupied by Turkey until recently. It's no wonder they prefer to remember their salad days, when their best thinkers came up with concepts like democracy and atoms. Greece is democratic again after a period of military dictatorship, and an atom is proudly pictured on one of their coins. That's why it hurts me to have to say that the ancient Greek hypothesis that matter is made of atoms was pure guesswork. There was no real experimental evidence for atoms, and the 18th-century revival of the atom concept by Dalton owed little to the Greeks other than the name, which means "unsplittable." Subtracting even more cruelly from Greek glory, the name was shown to be inappropriate in 1899 when physicist J.J. Thomson proved experimentally that atoms had even smaller things inside them, which could be extracted. (Thomson called them "electrons.") The "unsplittable" was splittable after all.

But that's getting ahead of our story. What happened to the atom concept in the intervening two thousand years? Educated people continued to discuss the idea, and those who were in favor of it could often use it to give plausible explanations for various facts and phenomena. One fact that was readily explained was conservation of mass. For example, if you mix 1 kg of water with 1 kg of dirt, you get exactly 2 kg of mud, no more and no less. The same is true for the a variety of processes such as freezing of water, fermenting beer, or pulverizing sandstone. If you believed in atoms, conservation of mass made perfect sense, because all these processes could be interpreted as mixing and rearranging atoms, without changing the total number of atoms. Still, this is nothing like a proof that atoms exist.

If atoms did exist, what types of atoms were there, and what distinguished the different types from each other? Was it their sizes, their shapes, their weights, or some other quality? The chasm between the ancient and modern atomisms becomes evident when we consider the wild speculations that existed on these issues until the present century. The ancients decided that there were four types of atoms, earth, water, air and fire; the most popular view was that they were distinguished by their shapes. Water atoms were spherical, hence water's ability to flow smoothly. Fire atoms had sharp points, which was why fire hurt when it touched one's skin. (There was no concept of temperature until thousands of years later.) The drastically different modern understanding of the structure of atoms was achieved in the course of the revolutionary decade stretching 1895 to 1905. The main purpose of this chapter is to describe those momentous experiments.

**Are you now or have you ever been an atomist?**

"You are what you eat." The glib modern phrase more or less assumes the atomic explanation of digestion. After all, digestion was pretty mysterious in ancient times, and premodern cultures would typically believe that eating allowed you to extract some kind of "life force" from the food. Myths abound to the effect that abstract qualities such as bravery or ritual impurity can enter your body via the food you eat. In contrast to these supernatural points of view, the ancient atomists had an entirely naturalistic interpretation of digestion. The food was made of atoms, and when you digested it you were simply extracting some atoms from it and rearranging them into the combina- tions required for your own body tissues. The more progressive medieval and renaissance scientists loved this kind of explanation. They were anxious to drive a stake through the heart of Aristotelian physics (and its embellished, Church-friendly version, scholasticism), which in their view ascribed too many occult properties and "purposes" to objects. For instance, the Aristotelian explanation for why a rock would fall to earth was that it was its "nature" or "purpose" to come to rest on the ground.

The seemingly innocent attempt to explain digestion naturalistically, however, ended up getting the atomists in big trouble with the Church. The problem was that the Church's most important sacrament involves eating bread and wine and thereby receiving the supernatural effect of forgiveness of sin. In connection with this ritual, the doctrine of transubstantiation asserts that the blessing of the eucharistic bread and wine literally transforms it into the blood and flesh of Christ. Atomism was perceived as contradicting transubstantiation, since atomism seemed to deny that the blessing could change the nature of the atoms. Although the historical information given in most science textbooks about Galileo represents his run-in with the Inquisition as turning on the issue of whether the earth moves, some historians believe his punishment had more to do with the perception that his advocacy of atomism subverted transubstantiation. (Other issues in the complex situation were Galileo's confrontational style, Pope Urban's military problems, and rumors that the stupid character in Galileo's dialogues was meant to be the Pope.) For a long time, belief in atomism served as a badge of nonconformity for scientists, a way of asserting a preference for natural rather than supernatural interpreta- tions of phenomena. Galileo and Newton's espousal of atomism was an act of rebellion, like later generations' adoption of Darwinism or Marxism.

Another conflict between scholasticism and atomism came from the question of what was between the atoms. If you ask modern people this question, they will probably reply "nothing" or "empty space." But Aristotle and his scholastic successors believed that there could be no such thing as empty space, i.e., a vacuum. That was not an unreasonable point of view, because air tends to rush in to any space you open up, and it wasn't until the renaissance that people figured out how to make a vacuum.

## Atoms, light, and everything else

Although I tend to ridicule ancient Greek philosophers like Aristotle, let's take a moment to praise him for something. If you read Aristotle's writings on physics (or just skim them, which is all I've done), the most striking thing is how careful he is about classifying phenomena and analyzing relationships among phenomena. The human brain seems to naturally make a distinction between two types of physical phenomena: objects and motion of objects. When a phenomenon occurs that does not immediately present itself as one of these, there is a strong tendency to conceptualize it as one or the other, or even to ignore its existence completely. For instance, physics teachers shudder at students' statements that "the dynamite exploded, and force came out of it in all directions." In these examples, the nonmaterial concept of force is being mentally categorized as if it was a physical substance. The statement that "winding the clock stores motion in the spring" is a miscategorization of electrical energy as a form of motion. An example of ignoring the existence of a phenomenon altogether can be elicited by asking people why we need lamps. The typical response that "the lamp illuminates the room so we can see things," ignores the necessary role of light coming into our eyes from the things being illuminated.

If you ask someone to tell you briefly about atoms, the likely response is that "everything is made of atoms," but we've now seen that it's far from obvious which "everything" this statement would properly refer to. For the scientists of the early 1900s who were trying to investigate atoms, this was not a trivial issue of definitions. There was a new gizmo called the vacuum tube, of which the only familiar example today is the picture tube of a TV. In short order, electrical tinkerers had discovered a whole flock of new phenomena that occurred in and around vacuum tubes, and given them picturesque names like "x-rays," "cathode rays," "Hertzian waves," and "N-rays." These were the types of observations that ended up telling us that we know about matter, but fierce controversies ensued over whether these were themselves forms of matter.

Let's bring ourselves up to the level of classification of phenomena employed by physicists in the year 1900. They recognized three categories:

- *Matter* has mass, can have kinetic energy, and can travel through a vacuum, transporting its mass and kinetic energy with it. Matter is conserved, both in the sense of conservation of mass and conservation of the number of atoms of each element. Atoms can't occupy the same space as other atoms, so a convenient way to prove something is not a form of matter is to show that it can pass through a solid material, in which the atoms are packed together closely.

- *Light* has no mass, always has energy, and can travel through a vacuum, transporting its energy with it. Two light beams can penetrate through each other and emerge from the collision without being weakened, deflected, or affected in any other way. Light can penetrate certain kinds of matter, e.g., glass.
- The third category is everything that doesn't fit the definition of light or matter. This catch-all category includes, for example, time, velocity, heat, and force.

### The chemical elements

How would one find out what types of atoms there were? Today, it doesn't seem like it should have been very difficult to work out an experimental program to classify the types of atoms. For each type of atom, there should be a corresponding element, i.e., a pure substance made out of nothing but that type of atom. Atoms are supposed to be unsplittable, so a substance like milk could not possibly be elemental, since churning it vigorously causes it to split up into two separate substances: butter and whey. Similarly, rust could not be an element, because it can be made by combining two substances: iron and oxygen. Despite its apparent reasonableness, no such program was carried out until the eighteenth century. The ancients presumably did not do it because observation was not universally agreed on as the right way to answer questions about nature, and also because they lacked the necessary techniques or the techniques were the province of laborers with low social status, such as smiths and miners. Alchemists were hindered by atomism's reputation for subversiveness, and by a tendency toward mysticism and secrecy. (The most celebrated challenge facing the alchemists, that of converting lead into gold, is one we now know to be impossible, since lead and gold are both elements.)

By 1900, however, chemists had done a reasonably good job of finding out what the elements were. They also had determined the ratios of the different atoms' masses fairly accurately. A typical technique would be to measure how many grams of sodium (Na) would combine with one gram of chlorine (Cl) to make salt (NaCl). (This assumes you've already decided based on other evidence that salt consisted of equal numbers of Na and Cl atoms.) The masses of individual atoms, as opposed to the mass ratios, were known only to within a few orders of magnitude based on indirect evidence, and plenty of physicists and chemists denied that individual atoms were anything more than convenient symbols.

$$\frac{m_{\mathrm{He}}}{m_{\mathrm{H}}} = 3.97$$
$$\frac{m_{\mathrm{Ne}}}{m_{\mathrm{H}}} = 20.01$$
$$\frac{m_{\mathrm{Sc}}}{m_{\mathrm{H}}} = 44.60$$

d / Examples of masses of atoms compared to that of hydrogen. Note how some, but not all, are close to integers.

### Making sense of the elements

As the information accumulated, the challenge was to find a way of systematizing it; the modern scientist's aesthetic sense rebels against complication. This hodgepodge of elements was an embarrassment. One contemporary observer, William Crookes, described the elements as extending “before us as stretched the wide Atlantic before the gaze of Columbus, mocking, taunting and murmuring strange riddles, which no man has yet been able to solve.” It wasn’t long before people started recognizing that many atoms’ masses were nearly integer multiples of the mass of hydrogen, the lightest element. A few excitable types began speculating that hydrogen was the basic building block, and that the heavier elements were made of clusters of hydrogen. It wasn’t long, however, before their parade was rained on by more accurate measurements, which showed that not all of the elements had atomic masses that were near integer multiples of hydrogen, and even the ones that were close to being integer multiples were off by one percent or so.

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 H | | | | | | | | | | | | | | | | | 2 He |
| 3 Li | 4 Be | | | | | | | | | | | 5 B | 6 C | 7 N | 8 O | 9 F | 10 Ne |
| 11 Na | 12 Mg | | | | | | | | | | | 13 Al | 14 Si | 15 P | 16 S | 17 Cl | 18 Ar |
| 19 K | 20 Ca | 21 Sc | 22 Ti | 23 V | 24 Cr | 25 Mn | 26 Fe | 27 Co | 28 Ni | 29 Cu | 30 Zn | 31 Ga | 32 Ge | 33 As | 34 Se | 35 Br | 36 Kr |
| 37 Rb | 38 Sr | 39 Y | 40 Zr | 41 Nb | 42 Mo | 43 Tc | 44 Ru | 45 Rh | 46 Pd | 47 Ag | 48 Cd | 49 In | 50 Sn | 51 Sb | 52 Te | 53 I | 54 Xe |
| 55 Cs | 56 Ba | 57 La * | 72 Hf | 73 Ta | 74 W | 75 Re | 76 Os | 77 Ir | 78 Pt | 79 Au | 80 Hg | 81 Tl | 82 Pb | 83 Bi | 84 Po | 85 At | 86 Rn |
| 87 Fr | 88 Ra | 89 Ac ** | 104 Rf | 105 Ha | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * | 58 Ce | 59 Pr | 60 Nd | 61 Pm | 62 Sm | 63 Eu | 64 Gd | 65 Tb | 66 Dy | 67 Ho | 68 Er | 69 Tm | 70 Yb | 71 Lu |
| ** | 90 Th | 91 Pa | 92 U | 93 Np | 94 Pu | 95 Am | 96 Cm | 97 Bk | 98 Cf | 99 Es | 100 Fm | 101 Md | 102 No | 103 Lr |

e / A modern periodic table. Elements in the same column have similar chemical properties. The modern atomic numbers, discussed in section 2.3, were not known in Mendeleev’s time, since the table could be flipped in various ways.

Chemistry professor Dmitri Mendeleev, preparing his lectures in 1869, wanted to find some way to organize his knowledge for his students to make it more understandable. He wrote the names of all the elements on cards and began arranging them in different ways on his desk, trying to find an arrangement that would make sense of the muddle. The row-and-column scheme he came up with is essentially our modern periodic table. The columns of the modern version represent groups of elements with similar chemical properties, and each row is more massive than the one above it. Going across each row, this almost always resulted in placing the atoms in sequence by weight as well. What made the system significant was its predictive value. There were three places where Mendeleev had to leave gaps in his checkerboard to keep chemically similar elements in the same column. He predicted that elements would exist to fill these gaps, and extrapolated or interpolated from other elements in the same column to predict their numerical properties, such as masses, boiling points, and densities. Mendeleev’s professional stock skyrocketed when his three elements (later named gallium, scandium and germanium) were discovered and found to have very nearly the properties he had predicted.

One thing that Mendeleev's table made clear was that mass was not the basic property that distinguished atoms of different elements. To make his table work, he had to deviate from ordering the elements strictly by mass. For instance, iodine atoms are lighter than tellurium, but Mendeleev had to put iodine after tellurium so that it would lie in a column with chemically similar elements.

### Direct proof that atoms existed

The success of the kinetic theory of heat was taken as strong evidence that, in addition to the motion of any object as a whole, there is an invisible type of motion all around us: the random motion of atoms within each object. But many conservatives were not convinced that atoms really existed. Nobody had ever seen one, after all. It wasn't until generations after the kinetic theory of heat was developed that it was demonstrated conclusively that atoms really existed and that they participated in continuous motion that never died out.

The smoking gun to prove atoms were more than mathematical abstractions came when some old, obscure observations were reexamined by an unknown Swiss patent clerk named Albert Einstein. A botanist named Brown, using a microscope that was state of the art in 1827, observed tiny grains of pollen in a drop of water on a microscope slide, and found that they jumped around randomly for no apparent reason. Wondering at first if the pollen he'd assumed to be dead was actually alive, he tried looking at particles of soot, and found that the soot particles also moved around. The same results would occur with any small grain or particle suspended in a liquid. The phenomenon came to be referred to as Brownian motion, and its existence was filed away as a quaint and thoroughly unimportant fact, really just a nuisance for the microscopist.

It wasn't until 1906 that Einstein found the correct interpretation for Brown's observation: the water molecules were in continuous random motion, and were colliding with the particle all the time, kicking it in random directions. After all the millennia of speculation about atoms, at last there was solid proof. Einstein's calculations dispelled all doubt, since he was able to make accurate predictions of things like the average distance traveled by the particle in a certain amount of time. (Einstein received the Nobel Prize not for his theory of relativity but for his papers on Brownian motion and the photoelectric effect.)

### Discussion Questions

**A** How could knowledge of the size of an individual aluminum atom be used to infer an estimate of its mass, or vice versa?

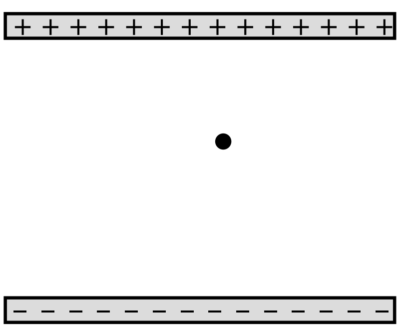
**B** How could one test Einstein's interpretation of Brownian motion by observing it at different temperatures?

## 1.4 Quantization of charge

Proving that atoms actually existed was a big accomplishment, but demonstrating their existence was different from understanding their properties. Note that the Brown-Einstein observations had nothing at all to do with electricity, and yet we know that matter is inherently electrical, and we have been successful in interpreting certain electrical phenomena in terms of mobile positively and negatively charged particles. Are these particles atoms? Parts of atoms? Particles that are entirely separate from atoms? It is perhaps premature to attempt to answer these questions without any conclusive evidence in favor of the charged-particle model of electricity.

f / A young Robert Millikan.

Strong support for the charged-particle model came from a 1911 experiment by physicist Robert Millikan at the University of Chicago. Consider a jet of droplets of perfume or some other liquid made by blowing it through a tiny pinhole. The droplets emerging from the pinhole must be smaller than the pinhole, and in fact most of them are even more microscopic than that, since the turbulent flow of air tends to break them up. Millikan reasoned that the droplets would acquire a little bit of electric charge as they rubbed against the channel through which they emerged, and if the charged-particle model of electricity was right, the charge might be split up among so many minuscule liquid drops that a single drop might have a total charge amounting to an excess of only a few charged particles — perhaps an excess of one positive particle on a certain drop, or an excess of two negative ones on another.



g / A simplified diagram of Millikan's apparatus.

Millikan's ingenious apparatus, g, consisted of two metal plates, which could be electrically charged as needed. He sprayed a cloud of oil droplets into the space between the plates, and selected one drop through a microscope for study. First, with no charge on the plates, he would determine the drop's mass by letting it fall through the air and measuring its terminal velocity, i.e., the velocity at which the force of air friction canceled out the force of gravity. The force of air drag on a slowly moving sphere had already been found by experiment to be $bvr^2$, where $b$ was a constant. Setting the total force equal to zero when the drop is at terminal velocity gives

$$bvr^2 - mg = 0 \quad ,$$

and setting the known density of oil equal to the drop's mass divided by its volume gives a second equation,

$$\rho = \frac{m}{\frac{4}{3}\pi r^3} \quad .$$

Everything in these equations can be measured directly except for $m$ and $r$, so these are two equations in two unknowns, which can be solved in order to determine how big the drop is.

Next Millikan charged the metal plates, adjusting the amount of charge so as to exactly counteract gravity and levitate the drop. If, for instance, the drop being examined happened to have a total charge that was negative, then positive charge put on the top plate would attract it, pulling it up, and negative charge on the bottom plate would repel it, pushing it up. (Theoretically only one plate would be necessary, but in practice a two-plate arrangement like this gave electrical forces that were more uniform in strength throughout the space where the oil drops were.) The amount of charge on the plates required to levitate the charged drop gave Millikan a handle on the amount of charge the drop carried. The more charge the drop had, the stronger the electrical forces on it would be, and the less charge would have to be put on the plates to do the trick. Unfortunately, expressing this relationship using Coulomb's law would have been impractical, because it would require a perfect knowledge of how the charge was distributed on each plate, plus the ability to perform vector addition of all the forces being exerted on the drop by all the charges on the plate. Instead, Millikan made use of the fact that the electrical force experienced by a pointlike charged object at a certain point in space is proportional to its charge,

$$\frac{F}{q} = \text{constant} \qquad .$$

With a given amount of charge on the plates, this constant could be determined for instance by discarding the oil drop, inserting between the plates a larger and more easily handled object with a known charge on it, and measuring the force with conventional methods. (Millikan actually used a slightly different set of techniques for determining the constant, but the concept is the same.) The amount of force on the actual oil drop had to equal $mg$, since it was just enough to levitate it, and once the calibration constant had been determined, the charge of the drop could then be found based on its previously determined mass.

| $q$ (C) | $q$ $/(1.64 \times 10^{-19}$ C) |
|---|---|
| $-1.970 \times 10^{-18}$ | $-12.02$ |
| $-0.987 \times 10^{-18}$ | $-6.02$ |
| $-2.773 \times 10^{-18}$ | $-16.93$ |

h / A few samples of Millikan's data.

Table h shows a few of the results from Millikan's 1911 paper. (Millikan took data on both negatively and positively charged drops, but in his paper he gave only a sample of his data on negatively charged drops, so these numbers are all negative.) Even a quick look at the data leads to the suspicion that the charges are not simply a series of random numbers. For instance, the second charge is almost exactly equal to half the first one. Millikan explained the observed charges as all being integer multiples of a single number, $1.64\times10^{-19}$ C. In the second column, dividing by this constant gives numbers that are essentially integers, allowing for the random errors present in the experiment. Millikan states in his paper that these results were a

> ... direct and tangible demonstration ... of the correctness of the view advanced many years ago and supported

> by evidence from many sources that all electrical charges, however produced, are exact multiples of one definite, elementary electrical charge, or in other words, that an electrical charge instead of being spread uniformly over the charged surface has a definite granular structure, consisting, in fact, of ...specks, or atoms of electricity, all precisely alike, peppered over the surface of the charged body.

In other words, he had provided direct evidence for the charged-particle model of electricity and against models in which electricity was described as some sort of fluid. The basic charge is notated $e$, and the modern value is $e = 1.60 \times 10^{-19}$ C. The word "*quantized*" is used in physics to describe a quantity that can only have certain numerical values, and cannot have any of the values between those. In this language, we would say that Millikan discovered that charge is quantized. The charge $e$ is referred to as the quantum of charge.

*self-check C*

Is money quantized? What is the quantum of money? ▷ Answer, p. 195

**A historical note on Millikan's fraud**

Very few undergraduate physics textbooks mention the well-documented fact that although Millikan's conclusions were correct, he was guilty of scientific fraud. His technique was difficult and painstaking to perform, and his original notebooks, which have been preserved, show that the data were far less perfect than he claimed in his published scientific papers. In his publications, he stated categorically that every single oil drop observed had had a charge that was a multiple of $e$, with no Exceptions or omissions. But his notebooks are replete with notations such as "beautiful data, keep," and "bad run, throw out." Millikan, then, appears to have earned his Nobel Prize by advocating a correct position with dishonest descriptions of his data.

Why do textbook authors fail to mention Millikan's fraud? It may be that they think students are too unsophisticated to correctly evaluate the implications of the fact that scientific fraud has sometimes existed and even been rewarded by the scientific establishment. Maybe they are afraid students will reason that fudging data is OK, since Millikan got the Nobel Prize for it. But falsifying history in the name of encouraging truthfulness is more than a little ironic. English teachers don't edit Shakespeare's tragedies so that the bad characters are always punished and the good ones never suffer!

Another possible explanation is simply a lack of originality; it's possible that some venerated textbook was uncritical of Millikan's fraud, and later authors simply followed suit. Biologist Stephen Jay Gould has written an essay tracing an example of how authors of biology textbooks tend to follow a certain traditional treatment of a topic, using the giraffe's neck to discuss the nonheritability of acquired traits. Yet another interpretation is that scientists derive status from their popular images as impartial searchers after the truth, and they don't want the public to realize how human and imperfect they can be. (Millikan himself was an educational reformer, and wrote a series of textbooks that were of much higher quality than others of his era.)

*Note added September 2002*
Several years after I wrote this historical digression, I came across an interesting defense of Millikan by David Goodstein (American Scientist, Jan-Feb 2001, pp. 54-60). Goodstein argues that although Millikan wrote a sentence in his paper that was a lie, Millikan is nevertheless not guilty of fraud when we take that sentence in context: Millikan stated that he never threw out any data, and he did throw out data, but he had good, objective reasons for throwing out the data. The Millikan affair will probably remain controversial among historians, but I would take away two lessons.

- The episode may reduce our confidence in Millikan, but it should deepen our faith in science. The correct result was eventually recognized; it might not have been in a pseudo-scientific field like medicine.
- In science, sloppiness can be almost as bad as cheating. If Science knows something of absolute truth, then she will not take excuses for falsehoods.

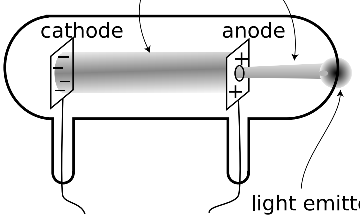
## 1.5 The electron

### Cathode rays

Nineteenth-century physicists spent a lot of time trying to come up with wild, random ways to play with electricity. The best experiments of this kind were the ones that made big sparks or pretty colors of light.

One such parlor trick was the cathode ray. To produce it, you first had to hire a good glassblower and find a good vacuum pump. The glassblower would create a hollow tube and embed two pieces of metal in it, called the electrodes, which were connected to the outside via metal wires passing through the glass. Before letting him seal up the whole tube, you would hook it up to a vacuum pump, and spend several hours huffing and puffing away at the pump's hand crank to get a good vacuum inside. Then, while you were still pumping on the tube, the glassblower would melt the glass and seal the whole thing shut. Finally, you would put a large amount of positive charge on one wire and a large amount of negative charge on the other. Metals have the property of letting charge move through them easily, so the charge deposited on one of the wires would quickly spread out because of the repulsion of each part of it for every other part. This spreading-out process would result in nearly all the charge ending up in the electrodes, where there is more room to spread out than there is in the wire. For obscure historical reasons a negative electrode is called a cathode and a positive one is an anode.

Figure i shows the light-emitting stream that was observed. If, as shown in this figure, a hole was made in the anode, the beam would extend on through the hole until it hit the glass. Drilling a hole in the cathode, however would not result in any beam coming out on the left side, and this indicated that the stuff, whatever it was, was coming from the cathode. The rays were therefore christened "cathode rays." (The terminology is still used today in the term "cathode ray tube" or "CRT" for the picture tube of a TV or computer monitor.)




i / Cathode rays observed in a vacuum tube.

## Were cathode rays a form of light, or of matter?

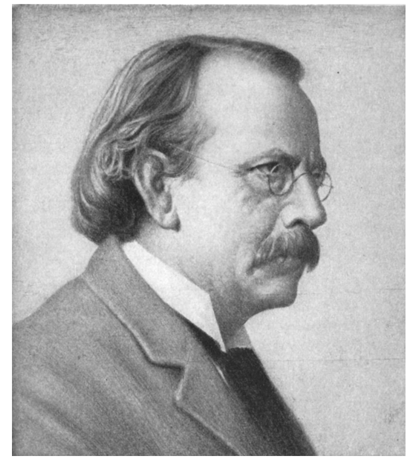
Were cathode rays a form of light, or matter? At first no one really cared what they were, but as their scientific importance became more apparent, the light-versus-matter issue turned into a controversy along nationalistic lines, with the Germans advocating light and the English holding out for matter. The supporters of the material interpretation imagined the rays as consisting of a stream of atoms ripped from the substance of the cathode.

One of our defining characteristics of matter is that material objects cannot pass through each other. Experiments showed that cathode rays could penetrate at least some small thickness of matter, such as a metal foil a tenth of a millimeter thick, implying that they were a form of light.

Other experiments, however, pointed to the contrary conclusion. Light is a wave phenomenon, and one distinguishing property of waves is demonstrated by speaking into one end of a paper towel roll. The sound waves do not emerge from the other end of the tube as a focused beam. Instead, they begin spreading out in all directions as soon as they emerge. This shows that waves do not necessarily travel in straight lines. If a piece of metal foil in the shape of a star or a cross was placed in the way of the cathode ray, then a "shadow" of the same shape would appear on the glass, showing that the rays traveled in straight lines. This straight-line motion suggested that they were a stream of small particles of matter.

These observations were inconclusive, so what was really needed was a determination of whether the rays had mass and weight. The trouble was that cathode rays could not simply be collected in a cup and put on a scale. When the cathode ray tube is in operation, one does not observe any loss of material from the cathode, or any crust being deposited on the anode.

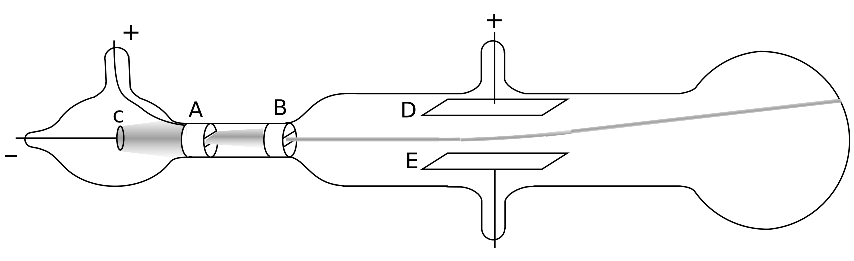
Nobody could think of a good way to weigh cathode rays, so the next most obvious way of settling the light/matter debate was to check whether the cathode rays possessed electrical charge. Light was known to be uncharged. If the cathode rays carried charge, they were definitely matter and not light, and they were presumably being made to jump the gap by the simultaneous repulsion of the negative charge in the cathode and attraction of the positive charge in the anode. The rays would overshoot the anode because of their momentum. (Although electrically charged particles do not normally leap across a gap of vacuum, very large amounts of charge were being used, so the forces were unusually intense.)



j / J.J. Thomson in the lab.

### Thomson's experiments

Physicist J.J. Thomson at Cambridge carried out a series of definitive experiments on cathode rays around the year 1897. By turning them slightly off course with electrical forces, k, he showed that they were indeed electrically charged, which was strong evidence that they were material. Not only that, but he proved that they had mass, and measured the ratio of their mass to their charge, $m/q$. Since their mass was not zero, he concluded that they were a form of matter, and presumably made up of a stream of microscopic, negatively charged particles. When Millikan published his results fourteen years later, it was reasonable to assume that the charge of one such particle equaled minus one fundamental charge, $q = -e$, and from the combination of Thomson's and Millikan's results one could therefore determine the mass of a single cathode ray particle.




k / Thomson's experiment proving cathode rays had electric charge (redrawn from his original paper). The cathode, c, and anode, A, are as in any cathode ray tube. The rays pass through a slit in the anode, and a second slit, B, is interposed in order to make the beam thinner and eliminate rays that were not going straight. Charging plates D and E shows that cathode rays have charge: they are attracted toward the positive plate D and repelled by the negative plate E.

The basic technique for determining $m/q$ was simply to measure the angle through which the charged plates bent the beam. The electric force acting on a cathode ray particle while it was between the plates would be proportional to its charge,

$$F_{elec} = (\text{known constant}) \cdot q \qquad .$$

Application of Newton's second law, $a = F/m$, would allow $m/q$ to be determined:

$$\frac{m}{q} = \frac{\text{known constant}}{a}$$

There was just one catch. Thomson needed to know the cathode ray particles' velocity in order to figure out their acceleration. At that point, however, nobody had even an educated guess as to the speed of the cathode rays produced in a given vacuum tube. The beam appeared to leap across the vacuum tube practically instantaneously, so it was no simple matter of timing it with a stopwatch!

Thomson's clever solution was to observe the effect of both electric and magnetic forces on the beam. The magnetic force exerted by a particular magnet would depend on both the cathode ray's charge and its speed:

$$F_{mag} = (\text{known constant \#2}) \cdot qv$$

Thomson played with the electric and magnetic forces until either one would produce an equal effect on the beam, allowing him to solve for the speed,

$$v = \frac{(\text{known constant})}{(\text{known constant \#2})} \quad .$$

Knowing the speed (which was on the order of 10% of the speed of light for his setup), he was able to find the acceleration and thus the mass-to-charge ratio $m/q$. Thomson's techniques were relatively crude (or perhaps more charitably we could say that they stretched the state of the art of the time), so with various methods he came up with $m/q$ values that ranged over about a factor of two, even for cathode rays extracted from a cathode made of a single material. The best modern value is $m/q = 5.69 \times 10^{-12}$ kg/C, which is consistent with the low end of Thomson's range.

### The cathode ray as a subatomic particle: the electron

What was significant about Thomson's experiment was not the actual numerical value of $m/q$, however, so much as the fact that, combined with Millikan's value of the fundamental charge, it gave a mass for the cathode ray particles that was thousands of times smaller than the mass of even the lightest atoms. Even without Millikan's results, which were 14 years in the future, Thomson recognized that the cathode rays' $m/q$ was thousands of times smaller than the $m/q$ ratios that had been measured for electrically charged atoms in chemical solutions. He correctly interpreted this as evidence that the cathode rays were smaller building blocks — he called them *electrons* — out of which atoms themselves were formed. This was an extremely radical claim, coming at a time when atoms had not yet been proven to exist! Even those who used the word "atom" often considered them no more than mathematical abstractions, not literal objects. The idea of searching for structure inside of "unsplittable" atoms was seen by some as lunacy, but within ten years Thomson's ideas had been amply verified by many more detailed experiments.

### Discussion Questions

**A** Thomson started to become convinced during his experiments that the "cathode rays" observed coming from the cathodes of vacuum tubes were building blocks of atoms — what we now call electrons. He then carried out observations with cathodes made of a variety of metals, and found that $m/q$ was roughly the same in every case, considering his limited accuracy. Given his suspicion, why did it make sense to try different metals? How would the consistent values of $m/q$ serve to test his hypothesis?

**B** My students have frequently asked whether the $m/q$ that Thomson measured was the value for a single electron, or for the whole beam. Can you answer this question?

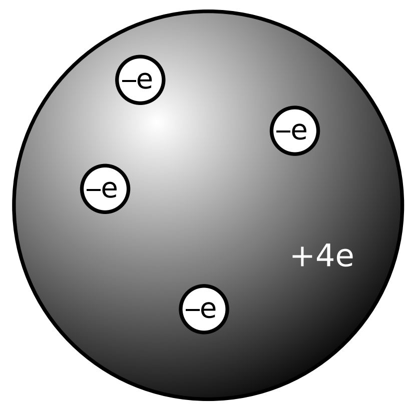
**C** Thomson found that the $m/q$ of an electron was thousands of times smaller than that of charged atoms in chemical solutions. Would this imply that the electrons had more charge? Less mass? Would there be no way to tell? Explain. Remember that Millikan's results were still many years in the future, so $q$ was unknown.

**D** Can you guess any practical reason why Thomson couldn't just let one electron fly across the gap before disconnecting the battery and turning off the beam, and then measure the amount of charge deposited on the anode, thus allowing him to measure the charge of a single electron directly?

**E** Why is it not possible to determine $m$ and $q$ themselves, rather than just their ratio, by observing electrons' motion in electric and magnetic fields?

## 1.6 The raisin cookie model of the atom

Based on his experiments, Thomson proposed a picture of the atom which became known as the raisin cookie model. In the neutral atom, l, there are four electrons with a total charge of $-4e$, sitting in a sphere (the "cookie") with a charge of $+4e$ spread throughout it. It was known that chemical reactions could not change one element into another, so in Thomson's scenario, each element's cookie sphere had a permanently fixed radius, mass, and positive charge, different from those of other elements. The electrons, however, were not a permanent feature of the atom, and could be tacked on or pulled out to make charged ions. Although we now know, for instance, that a neutral atom with four electrons is the element beryllium, scientists at the time did not know how many electrons the various neutral atoms possessed.



l / The raisin cookie model of the atom with four units of charge, which we now know to be beryllium.

This model is clearly different from the one you've learned in grade school or through popular culture, where the positive charge is concentrated in a tiny nucleus at the atom's center. An equally important change in ideas about the atom has been the realization that atoms and their constituent subatomic particles behave entirely differently from objects on the human scale. For instance, we'll see later that an electron can be in more than one place at one time. The raisin cookie model was part of a long tradition of attempts to make mechanical models of phenomena, and Thomson and his contemporaries never questioned the appropriateness of building a mental model of an atom as a machine with little parts inside. Today, mechanical models of atoms are still used (for instance the tinker-toy-style molecular modeling kits like the ones used by Watson and Crick to figure out the double helix structure of DNA), but scientists realize that the physical objects are only aids to help our brains' symbolic and visual processes think about atoms.

Although there was no clear-cut experimental evidence for many of the details of the raisin cookie model, physicists went ahead and started working out its implications. For instance, suppose you had a four-electron atom. All four electrons would be repelling each other, but they would also all be attracted toward the center of the "cookie" sphere. The result should be some kind of stable, symmetric arrangement in which all the forces canceled out. People sufficiently clever with math soon showed that the electrons in a four-electron atom should settle down at the vertices of a pyramid with one less side than the Egyptian kind, i.e., a regular tetrahedron. This deduction turns out to be wrong because it was based on incorrect features of the model, but the model also had many successes, a few of which we will now discuss.

*Flow of electrical charge in wires* *example 3*

One of my former students was the son of an electrician, and had become an electrician himself. He related to me how his father had remained refused to believe all his life that electrons really flowed through wires. If they had, he reasoned, the metal would have gradually become more and more damaged, eventually crumbling to dust.

His opinion is not at all unreasonable based on the fact that electrons are material particles, and that matter cannot normally pass through matter without making a hole through it. Nineteenth-century physicists would have shared his objection to a charged-particle model of the flow of electrical charge. In the raisin-cookie model, however, the electrons are very low in mass, and therefore presumably very small in size as well. It is not surprising that they can slip between the atoms without damaging them.

*Flow of electrical charge across cell membranes* *example 4*

Your nervous system is based on signals carried by charge moving from nerve cell to nerve cell. Your body is essentially all liquid, and atoms in a liquid are mobile. This means that, unlike the case of charge flowing in a solid wire, entire charged atoms can flow in your nervous system

*Emission of electrons in a cathode ray tube* *example 5*

Why do electrons detach themselves from the cathode of a vacuum tube? Certainly they are encouraged to do so by the repulsion of the negative charge placed on the cathode and the attraction from the net positive charge of the anode, but these are not strong enough to rip electrons out of atoms by main force — if they were, then the entire apparatus would have been instantly vaporized as every atom was simultaneously ripped apart!

The raisin cookie model leads to a simple explanation. We know that heat is the energy of random motion of atoms. The atoms in any object are therefore violently jostling each other all the time, and a few of these collisions are violent enough to knock electrons out of atoms. If this occurs near the surface of a solid object, the electron may can come loose. Ordinarily, however, this loss of electrons is a self-limiting process; the loss of electrons leaves the object with a net positive charge, which attracts the lost sheep home to the fold. (For objects immersed in air rather than vacuum, there will also be a balanced exchange of electrons between the air and the object.)

This interpretation explains the warm and friendly yellow glow of the vacuum tubes in an antique radio. To encourage the emission of electrons from the vacuum tubes' cathodes, the cathodes are intentionally warmed up with little heater coils.

### Discussion Questions

**A** Today many people would define an ion as an atom (or molecule) with missing electrons or extra electrons added on. How would people have defined the word "ion" before the discovery of the electron?

**B** Since electrically neutral atoms were known to exist, there had to be positively charged subatomic stuff to cancel out the negatively charged electrons in an atom. Based on the state of knowledge immediately after the Millikan and Thomson experiments, was it possible that the positively charged stuff had an unquantized amount of charge? Could it be quantized in units of +e? In units of +2e? In units of +5/7e?

## Summary

### Selected Vocabulary

| | |
|---|---|
| atom . . . . . . . | the basic unit of one of the chemical elements |
| molecule . . . . . | a group of atoms stuck together |
| electrical force . | one of the fundamental forces of nature; a non-contact force that can be either repulsive or attractive |
| charge . . . . . . | a numerical rating of how strongly an object participates in electrical forces |
| coulomb (C) the unit of electrical charge | an electrically charged atom or molecule |
| ion . . . . . . . . . | |
| cathode ray . . . | the mysterious ray that emanated from the cathode in a vacuum tube; shown by Thomson to be a stream of particles smaller than atoms |
| electron . . . . . . | Thomson's name for the particles of which a cathode ray was made |
| quantized . . . . | describes quantity such as money or electrical charge, that can only exist in certain amounts |

### Notation

| | |
|---|---|
| $q$ . . . . . . . . . . | charge |
| $e$ . . . . . . . . . . | the quantum of charge |

### Summary

All the forces we encounter in everyday life boil down to two basic types: gravitational forces and electrical forces. A force such as friction or a "sticky force" arises from electrical forces between individual atoms.

Just as we use the word "mass" to describe how strongly an object participates in gravitational forces, we use the word "charge" for the intensity of its electrical forces. There are two types of charge. Two charges of the same type repel each other, but objects whose charges are different attract each other. Charge is measured in units of coulombs (C).

Mobile charged particle model: A great many phenomena are easily understood if we imagine matter as containing two types of charged particles, which are at least partially able to move around.

Positive and negative charge: Ordinary objects that have not been specially prepared have both types of charge spread evenly throughout them in equal amounts. The object will then tend not to exert electrical forces on any other object, since any attraction due to one type of charge will be balanced by an equal repulsion from the other. (We say "tend not to" because bringing the object near an object with unbalanced amounts of charge could cause its charges to separate from each other, and the force would no longer cancel due to the unequal distances.) It therefore makes sense to describe the two types of charge using positive and negative signs, so that an unprepared object will have zero *total* charge.

The Coulomb force law states that the magnitude of the electrical force between two charged particles is given by $|\mathbf{F}| = k|q_1||q_2|/r^2$.

Conservation of charge: An even more fundamental reason for using positive and negative signs for charge is that with this definition the total charge of a closed system is a conserved quantity.

Quantization of charge: Millikan's oil drop experiment showed that the total charge of an object could only be an integer multiple of a basic unit of charge ($e$). This supported the idea the the "flow" of electrical charge was the motion of tiny particles rather than the motion of some sort of mysterious electrical fluid.

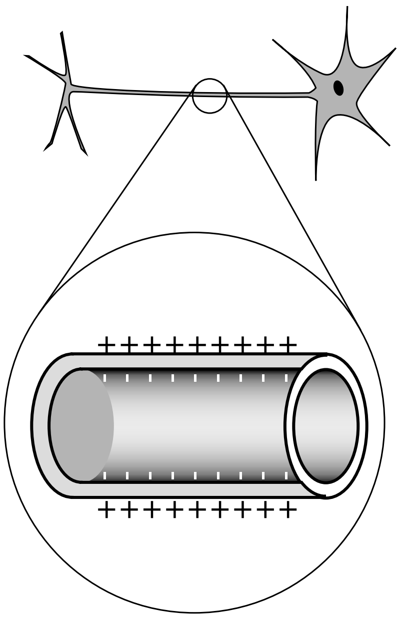
Einstein's analysis of Brownian motion was the first definitive proof of the existence of atoms. Thomson's experiments with vacuum tubes demonstrated the existence of a new type of microscopic particle with a very small ratio of mass to charge. Thomson correctly interpreted these as building blocks of matter even smaller than atoms: the first discovery of subatomic particles. These particles are called electrons.

The above experimental evidence led to the first useful model of the interior structure of atoms, called the raisin cookie model. In the raisin cookie model, an atom consists of a relatively large, massive, positively charged sphere with a certain number of negatively charged electrons embedded in it.

## Problems

### Key

- √ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ⋆ A difficult problem.




Problem 1. Top: A realistic picture of a neuron. Bottom: A simplified diagram of one segment of the tail (axon).

**1** The figure shows a neuron, which is the type of cell your nerves are made of. Neurons serve to transmit sensory information to the brain, and commands from the brain to the muscles. All this data is transmitted electrically, but even when the cell is resting and not transmitting any information, there is a layer of negative electrical charge on the inside of the cell membrane, and a layer of positive charge just outside it. This charge is in the form of various ions dissolved in the interior and exterior fluids. Why would the negative charge remain plastered against the inside surface of the membrane, and likewise why doesn't the positive charge wander away from the outside surface?

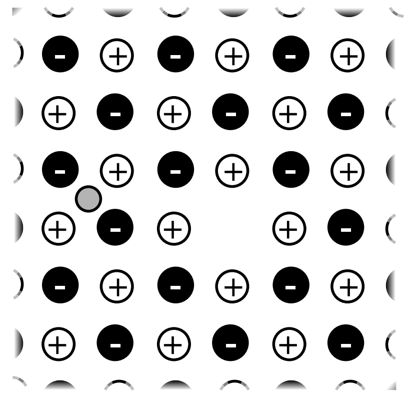
**2** Use the nutritional information on some packaged food to make an order-of-magnitude estimate of the amount of chemical energy stored in one atom of food, in units of joules. Assume that a typical atom has a mass of $10^{-26}$ kg. This constitutes a rough estimate of the amounts of energy there are on the atomic scale. [See chapter 1 of book 1, Newtonian Physics, for help on how to do order-of-magnitude estimates. Note that a nutritional "calorie" is really a kilocalorie; see page 202.] √

**3** (a) Recall that the gravitational energy of two gravitationally interacting spheres is given by $PE = -Gm_1m_2/r$, where $r$ is the center-to-center distance. What would be the analogous equation for two electrically interacting spheres? Justify your choice of a plus or minus sign on physical grounds, considering attraction and repulsion. √
(b) Use this expression to estimate the energy required to pull apart a raisin-cookie atom of the one-electron type, assuming a radius of $10^{-10}$ m. √
(c) Compare this with the result of problem 2.

**4** A neon light consists of a long glass tube full of neon, with metal caps on the ends. Positive charge is placed on one end of the tube, and negative charge on the other. The electric forces generated can be strong enough to strip electrons off of a certain number of neon atoms. Assume for simplicity that only one electron is ever stripped off of any neon atom. When an electron is stripped off of an atom, both the electron and the neon atom (now an ion) have electric charge, and they are accelerated by the forces exerted by the charged ends of the tube. (They do not feel any significant forces from the other ions and electrons within the tube, because only a tiny minority of neon atoms ever gets ionized.) Light is finally produced when ions are reunited with electrons. Give a numerical comparison of the magnitudes and directions of the accelerations of the electrons and ions. [You may need some data from page 202.] √

**5** If you put two hydrogen atoms near each other, they will feel an attractive force, and they will pull together to form a molecule. (Molecules consisting of two hydrogen atoms are the normal form of hydrogen gas.) Why do they feel a force if they are near each other, since each is electrically neutral? Shouldn't the attractive and repulsive forces all cancel out exactly? Use the raisin cookie model. (Students who have taken chemistry often try to use fancier models to explain this, but if you can't explain it using a simple model, you probably don't understand the fancy model as well as you thought you did!)

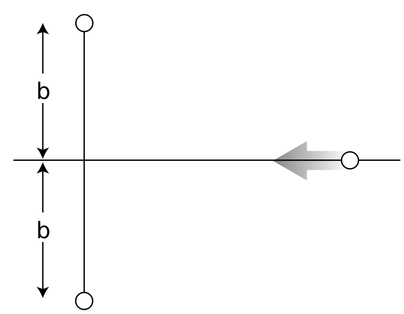
**6** The figure shows one layer of the three-dimensional structure of a salt crystal. The atoms extend much farther off in all directions, but only a six-by-six square is shown here. The larger circles are the chlorine ions, which have charges of $-e$. The smaller circles are sodium ions, with charges of $+e$. The center-to-center distance between neighboring ions is about 0.3 nm. Real crystals are never perfect, and the crystal shown here has two defects: a missing atom at one location, and an extra lithium atom, shown as a grey circle, inserted in one of the small gaps. If the lithium atom has a charge of $+e$, what is the direction and magnitude of the total force on it? Assume there are no other defects nearby in the crystal besides the two shown here. [Hints: The force on the lithium ion is the vector sum of all the forces of all the quadrillions of sodium and chlorine atoms, which would obviously be too laborious to calculate. Nearly all of these forces, however, are canceled by a force from an ion on the opposite side of the lithium.] √ ⋆



Problem 6.

**7** The Earth and Moon are bound together by gravity. If, instead, the force of attraction were the result of each having a charge of the same magnitude but opposite in sign, find the quantity of charge that would have to be placed on each to produce the required force. √

**8** In the semifinals of an electrostatic croquet tournament, Jessica hits her positively charged ball, sending it across the playing field, rolling to the left along the $x$ axis. It is repelled by two other positive charges. These two equal charges are fixed on the $y$ axis at the locations shown in the figure. (a) Express the force on the ball in terms of the ball's position, $x$. (b) At what value of $x$ does the ball experience the greatest deceleration? Express you answer in terms of $b$. [Based on a problem by Halliday and Resnick.] ∫




Problem 8.

a / Marie and Pierre Curie were the first to purify radium in significant quantities. Radium's intense radioactivity made possible the experiments that led to the modern planetary model of the atom, in which electrons orbit a nucleus made of protons and neutrons.

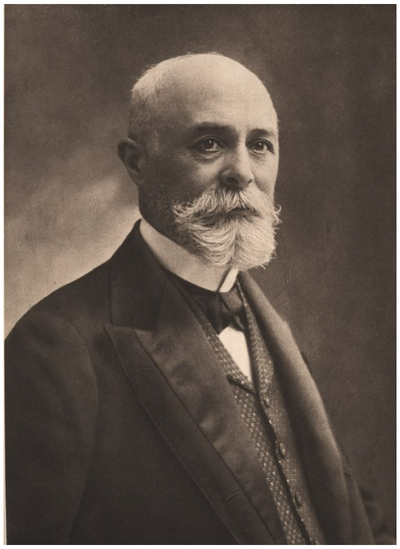
# Chapter 2
# The Nucleus

## 2.1 Radioactivity

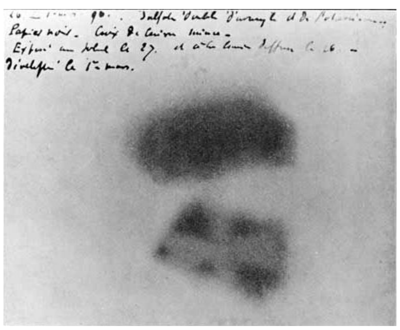
### Becquerel's discovery of radioactivity

How did physicists figure out that the raisin cookie model was incorrect, and that the atom's positive charge was concentrated in a tiny, central nucleus? The story begins with the discovery of radioactivity by the French chemist Becquerel. Up until radioactivity was discovered, all the processes of nature were thought to be based on chemical reactions, which were rearrangements of combinations of atoms. Atoms exert forces on each other when they are close together, so sticking or unsticking them would either release or store electrical energy. That energy could be converted to and from other forms, as when a plant uses the energy in sunlight to make sugars and carbohydrates, or when a child eats sugar, releasing the energy in the form of kinetic energy.

Becquerel discovered a process that seemed to release energy from an unknown new source that was not chemical. Becquerel,



b / Henri Becquerel (1852-1908).



c / Becquerel's photographic plate. In the exposure at the bottom of the image, he has found that he could absorb the radiations, casting the shadow of a Maltese cross that was placed between the plate and the uranium salts.

whose father and grandfather had also been physicists, spent the first twenty years of his professional life as a successful civil engineer, teaching physics on a part-time basis. He was awarded the chair of physics at the Musée d'Histoire Naturelle in Paris after the death of his father, who had previously occupied it. Having now a significant amount of time to devote to physics, he began studying the interaction of light and matter. He became interested in the phenomenon of phosphorescence, in which a substance absorbs energy from light, then releases the energy via a glow that only gradually goes away. One of the substances he investigated was a uranium compound, the salt $UKSO_5$. One day in 1896, cloudy weather interfered with his plan to expose this substance to sunlight in order to observe its fluorescence. He stuck it in a drawer, coincidentally on top of a blank photographic plate — the old-fashioned glass-backed counterpart of the modern plastic roll of film. The plate had been carefully wrapped, but several days later when Becquerel checked it in the darkroom before using it, he found that it was ruined, as if it had been completely exposed to light.

History provides many examples of scientific discoveries that happened this way: an alert and inquisitive mind decides to investigate a phenomenon that most people would not have worried about explaining. Becquerel first determined by further experiments that the effect was produced by the uranium salt, despite a thick wrapping of paper around the plate that blocked out all light. He tried a variety of compounds, and found that it was the uranium that did it: the effect was produced by any uranium compound, but not by any compound that didn't include uranium atoms. The effect could be at least partially blocked by a sufficient thickness of metal, and he was able to produce silhouettes of coins by interposing them between the uranium and the plate. This indicated that the effect traveled in a straight line., so that it must have been some kind of ray rather than, e.g., the seepage of chemicals through the paper. He used the word "radiations," since the effect radiated out from the uranium salt.

At this point Becquerel still believed that the uranium atoms were absorbing energy from light and then gradually releasing the energy in the form of the mysterious rays, and this was how he presented it in his first published lecture describing his experiments. Interesting, but not earth-shattering. But he then tried to determine how long it took for the uranium to use up all the energy that had supposedly been stored in it by light, and he found that it never seemed to become inactive, no matter how long he waited. Not only that, but a sample that had been exposed to intense sunlight for a whole afternoon was no more or less effective than a sample that had always been kept inside. Was this a violation of conservation of energy? If the energy didn't come from exposure to light, where did it come from?

### Three kinds of "radiations"

Unable to determine the source of the energy directly, turn-of-the-century physicists instead studied the behavior of the "radiations" once they had been emitted. Becquerel had already shown that the radioactivity could penetrate through cloth and paper, so the first obvious thing to do was to investigate in more detail what thickness of material the radioactivity could get through. They soon learned that a certain fraction of the radioactivity's intensity would be eliminated by even a few inches of air, but the remainder was not eliminated by passing through more air. Apparently, then, the radioactivity was a mixture of more than one type, of which one was blocked by air. They then found that of the part that could penetrate air, a further fraction could be eliminated by a piece of paper or a very thin metal foil. What was left after that, however, was a third, extremely penetrating type, some of whose intensity would still remain even after passing through a brick wall. They decided that this showed there were three types of radioactivity, and without having the faintest idea of what they really were, they made up names for them. The least penetrating type was arbitrarily labeled $\alpha$ (alpha), the first letter of the Greek alphabet, and so on through $\beta$ (beta) and finally $\gamma$ (gamma) for the most penetrating type.

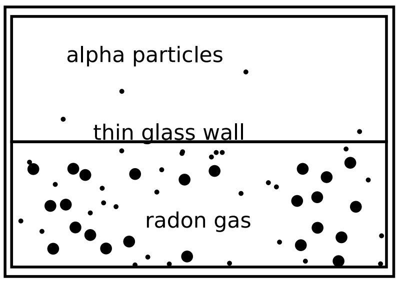
### Radium: a more intense source of radioactivity

The measuring devices used to detect radioactivity were crude: photographic plates or even human eyeballs (radioactivity makes flashes of light in the jelly-like fluid inside the eye, which can be seen by the eyeball's owner if it is otherwise very dark). Because the ways of detecting radioactivity were so crude and insensitive, further progress was hindered by the fact that the amount of radioactivity emitted by uranium was not really very great. The vital contribution of physicist/chemist Marie Curie and her husband Pierre was to discover the element radium, and to purify and isolate significant quantities it. Radium emits about a million times more radioactivity per unit mass than uranium, making it possible to do the experiments that were needed to learn the true nature of radioactivity. The dangers of radioactivity to human health were then unknown, and Marie died of leukemia thirty years later. (Pierre was run over and killed by a horsecart.)

### Tracking down the nature of alphas, betas, and gammas

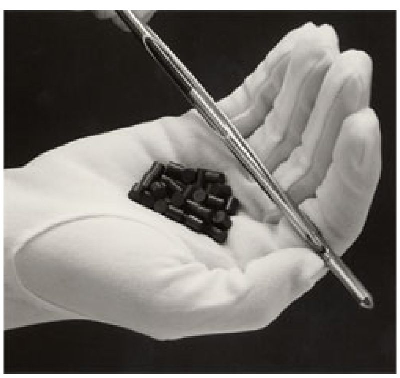
As radium was becoming available, an apprentice scientist named Ernest Rutherford arrived in England from his native New Zealand and began studying radioactivity at the Cavendish Laboratory. The young colonial's first success was to measure the mass-to-charge ratio of beta rays. The technique was essentially the same as the one Thomson had used to measure the mass-to-charge ratio of cathode rays by measuring their deflections in electric and magnetic fields. The only difference was that instead of the cathode of a vacuum tube, a nugget of radium was used to supply the beta rays. Not only was the technique the same, but so was the result. Beta rays had the same $m/q$ ratio as cathode rays, which suggested they were one and the same. Nowadays, it would make sense simply to use the term "electron," and avoid the archaic "cathode ray" and "beta particle," but the old labels are still widely used, and it is unfortunately necessary for physics students to memorize all three names for the same thing.

At first, it seemed that neither alphas or gammas could be deflected in electric or magnetic fields, making it appear that neither was electrically charged. But soon Rutherford obtained a much more powerful magnet, and was able to use it to deflect the alphas but not the gammas. The alphas had a much larger value of $m/q$ than the betas (about 4000 times greater), which was why they had been so hard to deflect. Gammas are uncharged, and were later found to be a form of light.




d / A simplified version of Rutherford's 1908 experiment, showing that alpha particles were doubly ionized helium atoms.

The $m/q$ ratio of alpha particles turned out to be the same as those of two different types of ions, $He^{++}$ (a helium atom with two missing electrons) and $H_2^+$ (two hydrogen atoms bonded into a molecule, with one electron missing), so it seemed likely that they were one or the other of those. The diagram shows a simplified version of Rutherford's ingenious experiment proving that they were $He^{++}$ ions. The gaseous element radon, an alpha emitter, was introduced into one half of a double glass chamber. The glass wall dividing the chamber was made extremely thin, so that some of the rapidly moving alpha particles were able to penetrate it. The other chamber, which was initially evacuated, gradually began to accumulate a population of alpha particles (which would quickly pick up electrons from their surroundings and become electrically neutral). Rutherford then determined that it was helium gas that had appeared in the second chamber. Thus alpha particles were proved to be $He^{++}$ ions. The nucleus was yet to be discovered, but in modern terms, we would describe a $He^{++}$ ion as the nucleus of a He atom.



e / These pellets of uranium fuel will be inserted into the metal fuel rod and used in a nuclear reactor. The pellets emit alpha and beta radiation, which the gloves are thick enough to stop.

To summarize, here are the three types of radiation emitted by radioactive elements, and their descriptions in modern terms:

| | | |
|---|---|---|
| $\alpha$ particle | stopped by a few inches of air | He nucleus |
| $\beta$ particle | stopped by a piece of paper | electron |
| $\gamma$ ray | penetrates thick shielding | a type of light |

### Discussion Questions

**A** Most sources of radioactivity emit alphas, betas, and gammas, not just one of the three. In the radon experiment, how did Rutherford know that he was studying the alphas?

## 2.2 The planetary model of the atom

The stage was now set for the unexpected discovery that the positively charged part of the atom was a tiny, dense lump at the atom's center rather than the "cookie dough" of the raisin cookie model. By 1909, Rutherford was an established professor, and had students working under him. For a raw undergraduate named Marsden, he picked a research project he thought would be tedious but straightforward.

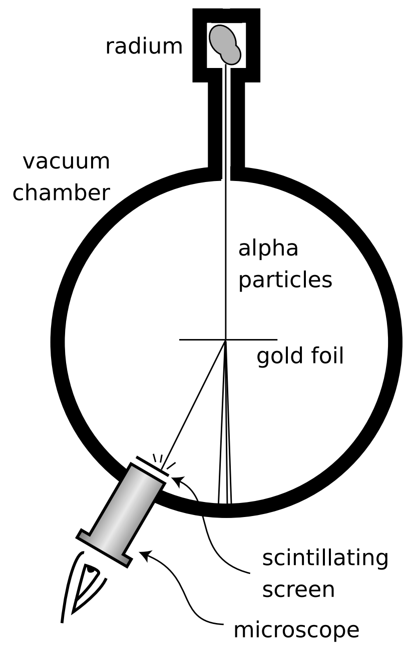
It was already known that although alpha particles would be stopped completely by a sheet of paper, they could pass through a sufficiently thin metal foil. Marsden was to work with a gold foil only 1000 atoms thick. (The foil was probably made by evaporating a little gold in a vacuum chamber so that a thin layer would be deposited on a glass microscope slide. The foil would then be lifted off the slide by submerging the slide in water.)

Rutherford had already determined in his previous experiments the speed of the alpha particles emitted by radium, a fantastic $1.5 \times 10^7$ m/s. The experimenters in Rutherford's group visualized them as very small, very fast cannonballs penetrating the "cookie dough" part of the big gold atoms. A piece of paper has a thickness of a hundred thousand atoms or so, which would be sufficient to stop them completely, but crashing through a thousand would only slow them a little and turn them slightly off of their original paths.

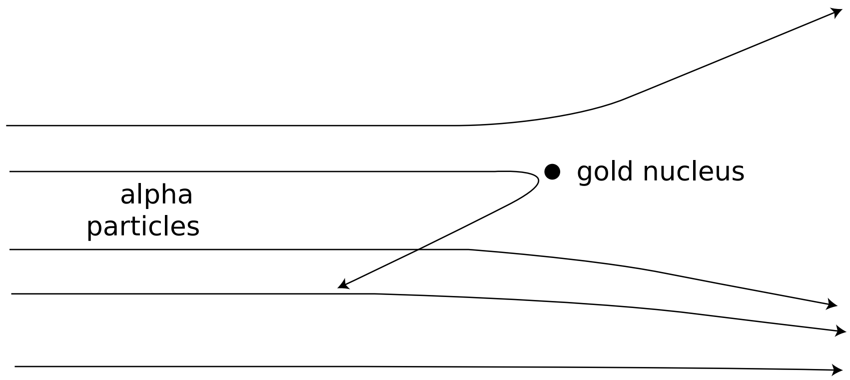
Marsden's supposedly ho-hum assignment was to use the apparatus shown in figure g to measure how often alpha particles were deflected at various angles. A tiny lump of radium in a box emitted alpha particles, and a thin beam was created by blocking all the alphas except those that happened to pass out through a tube. Typically deflected in the gold by only a small amount, they would reach a screen very much like the screen of a TV's picture tube, which would make a flash of light when it was hit. Here is the first example we have encountered of an experiment in which a beam of particles is detected one at a time. This was possible because each alpha particle carried so much kinetic energy; they were moving at about the same speed as the electrons in the Thomson experiment, but had ten thousand times more mass.

Marsden sat in a dark room, watching the apparatus hour after hour and recording the number of flashes with the screen moved to various angles. The rate of the flashes was highest when he set the screen at an angle close to the line of the alphas' original path, but if he watched an area farther off to the side, he would also occasionally see an alpha that had been deflected through a larger angle. After seeing a few of these, he got the crazy idea of moving the screen to see if even larger angles ever occurred, perhaps even angles larger than 90 degrees.

f / Ernest Rutherford (1871-1937).

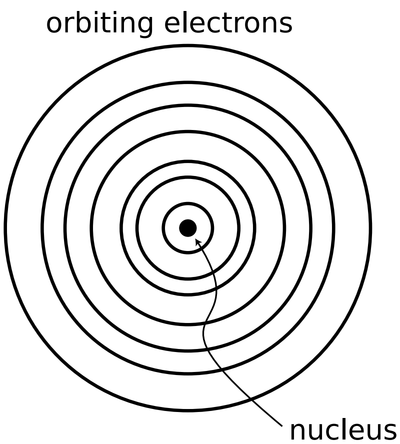



g / Marsden and Rutherford's apparatus.




h / Alpha particles being scattered by a gold nucleus. On this scale, the gold atom is the size of a car, so all the alpha particles shown here are ones that just happened to come unusually close to the nucleus. For these exceptional alpha particles, the forces from the electrons are unimportant, because they are so much more distant than the nucleus.

The crazy idea worked: a few alpha particles were deflected through angles of up to 180 degrees, and the routine experiment had become an epoch-making one. Rutherford said, “We have been able to get some of the alpha particles coming backwards. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.” Explanations were hard to come by in the raisin cookie model. What intense electrical forces could have caused some of the alpha particles, moving at such astronomical speeds, to change direction so drastically? Since each gold atom was electrically neutral, it would not exert much force on an alpha particle outside it. True, if the alpha particle was very near to or inside of a particular atom, then the forces would not necessarily cancel out perfectly; if the alpha particle happened to come very close to a particular electron, the $1/r^2$ form of the Coulomb force law would make for a very strong force. But Marsden and Rutherford knew that an alpha particle was 8000 times more massive than an electron, and it is simply not possible for a more massive object to rebound backwards from a collision with a less massive object while conserving momentum and energy. It might be possible in principle for a particular alpha to follow a path that took it very close to one electron, and then very close to another electron, and so on, with the net result of a large deflection, but careful calculations showed that such multiple “close encounters” with electrons would be millions of times too rare to explain what was actually observed.




i / The planetary model of the atom.

At this point, Rutherford and Marsden dusted off an unpopular and neglected model of the atom, in which all the electrons orbited around a small, positively charged core or “nucleus,” just like the planets orbiting around the sun. All the positive charge and nearly all the mass of the atom would be concentrated in the nucleus, rather than spread throughout the atom as in the raisin cookie model. The positively charged alpha particles would be repelled by the gold atom's nucleus, but most of the alphas would not come close enough to any nucleus to have their paths drastically altered. The few that did come close to a nucleus, however, could rebound backwards from a single such encounter, since the nucleus of a heavy gold atom would be fifty times more massive than an alpha particle. It turned out that it was not even too difficult to derive a formula giving the relative frequency of deflections through various angles, and this calculation agreed with the data well enough (to within 15%), considering the difficulty in getting good experimental statistics on the rare, very large angles.

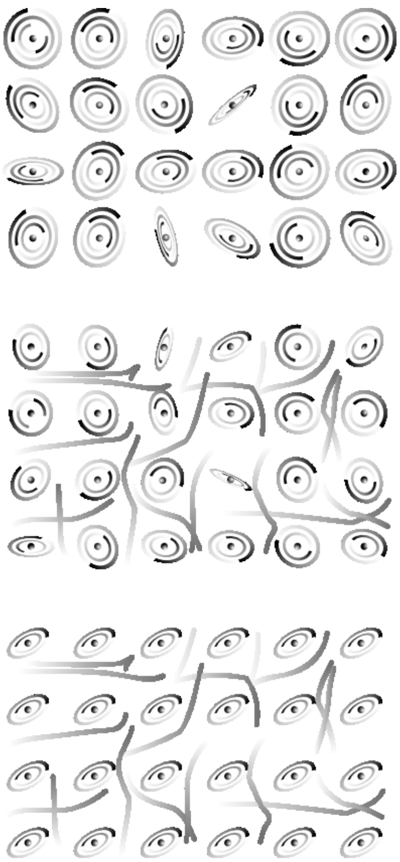
What had started out as a tedious exercise to get a student started in science had ended as a revolution in our understanding of nature. Indeed, the whole thing may sound a little too much like a moralistic fable of the scientific method with overtones of the Horatio Alger genre. The skeptical reader may wonder why the planetary model was ignored so thoroughly until Marsden and Rutherford's discovery. Is science really more of a sociological enterprise, in which certain ideas become accepted by the establishment, and other, equally plausible explanations are arbitrarily discarded? Some social scientists are currently ruffling a lot of scientists' feathers with critiques very much like this, but in this particular case, there were very sound reasons for rejecting the planetary model. As you'll learn in more detail later in this course, any charged particle that undergoes an acceleration dissipate energy in the form of light. In the planetary model, the electrons were orbiting the nucleus in circles or ellipses, which meant they were undergoing acceleration, just like the acceleration you feel in a car going around a curve. They should have dissipated energy as light, and eventually they should have lost all their energy. Atoms don't spontaneously collapse like that, which was why the raisin cookie model, with its stationary electrons, was originally preferred. There were other problems as well. In the planetary model, the one-electron atom would have to be flat, which would be inconsistent with the success of molecular modeling with spherical balls representing hydrogen and atoms. These molecular models also seemed to work best if specific sizes were used for different atoms, but there is no obvious reason in the planetary model why the radius of an electron's orbit should be a fixed number. In view of the conclusive Marsden-Rutherford results, however, these became fresh puzzles in atomic physics, not reasons for disbelieving the planetary model.

### Some phenomena explained with the planetary model

The planetary model may not be the ultimate, perfect model of the atom, but don't underestimate its power. It already allows us to visualize correctly a great many phenomena.

As an example, let's consider the distinctions among nonmetals, metals that are magnetic, and metals that are nonmagnetic. As shown in figure j, a metal differs from a nonmetal because its outermost electrons are free to wander rather than owing their allegiance to a particular atom. A metal that can be magnetized is one that is willing to line up the rotations of some of its electrons so that their axes are parallel. Recall that magnetic forces are forces made by moving charges; we have not yet discussed the mathematics and geometry of magnetic forces, but it is easy to see how random orientations of the atoms in the nonmagnetic substance would lead to cancellation of the forces.

Even if the planetary model does not immediately answer such questions as why one element would be a metal and another a nonmetal, these ideas would be difficult or impossible to conceptualize in the raisin cookie model.



j / The planetary model applied to a nonmetal, 1, an unmagnetized metal, 2, and a magnetized metal, 3. Note that these figures are all simplified in several ways. For one thing, the electrons of an individual atom do not all revolve around the nucleus in the same plane. It is also very unusual for a metal to become so strongly magnetized that 100% of its atoms have their rotations aligned as shown in this figure.

### Discussion Questions

**A** In reality, charges of the same type repel one another and charges of different types are attracted. Suppose the rules were the other way around, giving repulsion between opposite charges and attraction between similar ones. What would the universe be like?

## 2.3 Atomic number

As alluded to in a discussion question in the previous section, scientists of this period had only a very approximate idea of how many units of charge resided in the nuclei of the various chemical elements. Although we now associate the number of units of nuclear charge with the element's position on the periodic table, and call it the atomic number, they had no idea that such a relationship existed. Mendeleev's table just seemed like an organizational tool, not something with any necessary physical significance. And everything Mendeleev had done seemed equally valid if you turned the table upside-down or reversed its left and right sides, so even if you wanted to number the elements sequentially with integers, there was an ambiguity as to how to do it. Mendeleev's original table was in fact upside-down compared to the modern one.

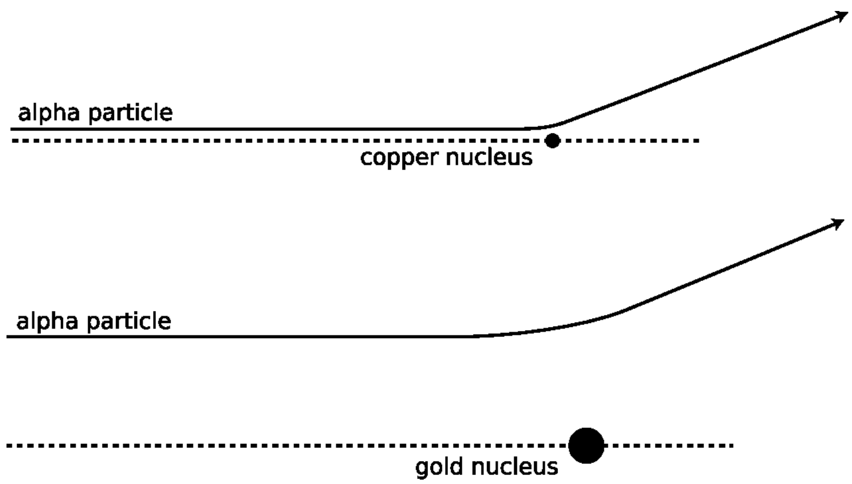
| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 H | | | | | | | | | | | | | | | | | 2 He |
| 3 Li | 4 Be | | | | | | | | | | | 5 B | 6 C | 7 N | 8 O | 9 F | 10 Ne |
| 11 Na | 12 Mg | | | | | | | | | | | 13 Al | 14 Si | 15 P | 16 S | 17 Cl | 18 Ar |
| 19 K | 20 Ca | 21 Sc | 22 Ti | 23 V | 24 Cr | 25 Mn | 26 Fe | 27 Co | 28 Ni | 29 Cu | 30 Zn | 31 Ga | 32 Ge | 33 As | 34 Se | 35 Br | 36 Kr |
| 37 Rb | 38 Sr | 39 Y | 40 Zr | 41 Nb | 42 Mo | 43 Tc | 44 Ru | 45 Rh | 46 Pd | 47 Ag | 48 Cd | 49 In | 50 Sn | 51 Sb | 52 Te | 53 I | 54 Xe |
| 55 Cs | 56 Ba | 57 La * | 72 Hf | 73 Ta | 74 W | 75 Re | 76 Os | 77 Ir | 78 Pt | 79 Au | 80 Hg | 81 Tl | 82 Pb | 83 Bi | 84 Po | 85 At | 86 Rn |
| 87 Fr | 88 Ra | 89 Ac ** | 104 Rf | 105 Ha | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * | 58 Ce | 59 Pr | 60 Nd | 61 Pm | 62 Sm | 63 Eu | 64 Gd | 65 Tb | 66 Dy | 67 Ho | 68 Er | 69 Tm | 70 Yb | 71 Lu |
| ** | 90 Th | 91 Pa | 92 U | 93 Np | 94 Pu | 95 Am | 96 Cm | 97 Bk | 98 Cf | 99 Es | 100 Fm | 101 Md | 102 No | 103 Lr |

k / A modern periodic table, labeled with atomic numbers. Mendeleev's original table was upside-down compared to this one.

In the period immediately following the discovery of the nucleus, physicists only had rough estimates of the charges of the various nuclei. In the case of the very lightest nuclei, they simply found the maximum number of electrons they could strip off by various methods: chemical reactions, electric sparks, ultraviolet light, and so on. For example they could easily strip of one or two electrons from helium, making $He^+$ or $He^{++}$, but nobody could make $He^{+++}$, presumably because the nuclear charge of helium was only $+2e$. Unfortunately only a few of the lightest elements could be stripped completely, because the more electrons were stripped off, the greater the positive net charge remaining, and the more strongly the rest of the negatively charged electrons would be held on. The heavy elements' atomic numbers could only be roughly extrapolated from the light elements, where the atomic number was about half the atom's mass expressed in units of the mass of a hydrogen atom. Gold, for example, had a mass about 197 times that of hydrogen, so its atomic number was estimated to be about half that, or somewhere around 100. We now know it to be 79.

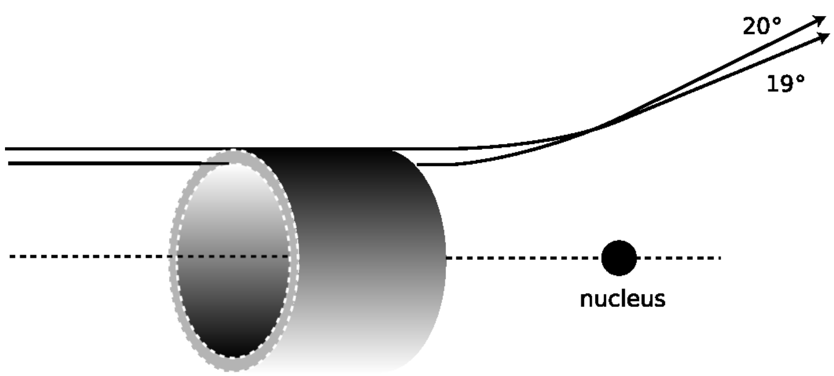
How did we finally find out? The riddle of the nuclear charges was at last successfully attacked using two different techniques, which gave consistent results. One set of experiments, involving x-rays, was performed by the young Henry Mosely, whose scientific brilliance was soon to be sacrificed in a battle between European imperialists over who would own the Dardanelles, during that pointless conflict then known as the War to End All Wars, and now referred to as World War I.




l / An alpha particle has to come much closer to the low-charged copper nucleus in order to be deflected through the same angle.

Since Mosely's analysis requires several concepts with which you are not yet familiar, we will instead describe the technique used by James Chadwick at around the same time. An added bonus of describing Chadwick's experiments is that they presaged the important modern technique of studying *collisions* of subatomic particles. In grad school, I worked with a professor whose thesis adviser's thesis adviser was Chadwick, and he related some interesting stories about the man. Chadwick was apparently a little nutty and a complete fanatic about science, to the extent that when he was held in a German prison camp during World War II, he managed to cajole his captors into allowing him to scrounge up parts from broken radios so that he could attempt to do physics experiments.

Chadwick's experiment worked like this. Suppose you perform two Rutherford-type alpha scattering measurements, first one with a gold foil as a target as in Rutherford's original experiment, and then one with a copper foil. It is possible to get large angles of deflection in both cases, but as shown in figure m, the alpha particle must be heading almost straight for the copper nucleus to get the same angle of deflection that would have occurred with an alpha that was much farther off the mark; the gold nucleus' charge is so much greater than the copper's that it exerts a strong force on the alpha particle even from far off. The situation is very much like that of a blindfolded person playing darts. Just as it is impossible to aim an alpha particle at an individual nucleus in the target, the blindfolded person cannot really aim the darts. Achieving a very close encounter with the copper atom would be akin to hitting an inner circle on the dartboard. It's much more likely that one would have the luck to hit the outer circle, which covers a greater number of square inches. By analogy, if you measure the frequency with which alphas are scattered by copper at some particular angle, say between 19 and 20 degrees, and then perform the same measurement at the same angle with gold, you get a much higher percentage for gold than for copper.




m / An alpha particle must be headed for the ring on the front of the imaginary cylindrical pipe in order to produce scattering at an angle between 19 and 20 degrees. The area of this ring is called the "cross-section" for scattering at 19-20° because it is the cross-sectional area of a cut through the pipe.

In fact, the numerical ratio of the two nuclei's charges can be derived from this same experimentally determined ratio. Using the standard notation $Z$ for the atomic number (charge of the nucleus divided by $e$), the following equation can be proved (example 1):

$$\frac{Z^2_{gold}}{Z^2_{copper}} = \frac{\text{number of alphas scattered by gold at 19-20}^\circ}{\text{number of alphas scattered by copper at 19-20}^\circ}$$

By making such measurements for targets constructed from all the elements, one can infer the ratios of all the atomic numbers, and since the atomic numbers of the light elements were already known, atomic numbers could be assigned to the entire periodic table. According to Mosely, the atomic numbers of copper, silver and platinum were 29, 47, and 78, which corresponded well with their positions on the periodic table. Chadwick's figures for the same elements were 29.3, 46.3, and 77.4, with error bars of about 1.5 times the fundamental charge, so the two experiments were in good agreement.

The point here is absolutely not that you should be ready to plug numbers into the above equation for a homework or exam question! My overall goal in this chapter is to explain how we know what we know about atoms. An added bonus of describing Chadwick's experiment is that the approach is very similar to that used in modern particle physics experiments, and the ideas used in the analysis are closely related to the now-ubiquitous concept of a "cross-section."

In the dartboard analogy, the cross-section would be the area of the circular ring you have to hit. The reasoning behind the invention of the term "cross-section" can be visualized as shown in figure m. In this language, Rutherford's invention of the planetary model came from his unexpected discovery that there was a nonzero cross-section for alpha scattering from gold at large angles, and Chadwick confirmed Mosely's determinations of the atomic numbers by measuring cross-sections for alpha scattering.

*Proof of the relationship between Z and scattering* *example 1*

The equation above can be derived by the following not very rigorous proof. To deflect the alpha particle by a certain angle requires that it acquire a certain momentum component in the direction perpendicular to its original momentum. Although the nucleus's force on the alpha particle is not constant, we can pretend that it is approximately constant during the time when the alpha is within a distance equal to, say, 150% of its distance of closest approach, and that the force is zero before and after that part of the motion. (If we chose 120% or 200%, it shouldn't make any difference in the final result, because the final result is a ratio, and the effects on the numerator and denominator should cancel each other.) In the approximation of constant force, the change in the alpha's perpendicular momentum component is then equal to $F\Delta t$. The Coulomb force law says the force is proportional to $Z/r^2$. Although $r$ does change somewhat during the time interval of interest, it's good enough to treat it as a constant number, since we're only computing the ratio between the two experiments' results. Since we are approximating the force as acting over the time during which the distance is not too much greater than the distance of closest approach, the time interval $\Delta t$ must be proportional to $r$, and the sideways momentum imparted to the alpha, $F\Delta t$, is proportional to $(Z/r^2)r$, or $Z/r$. If we're comparing alphas scattered at the same angle from gold and from copper, then $\Delta p$ is the same in both cases, and the proportionality $\Delta p \propto Z/r$ tells us that the ones scattered from copper at that angle had to be headed in along a line closer to the central axis by a factor equaling $Z_{\text{gold}}/Z_{\text{copper}}$. If you imagine a "dartboard ring" that the alphas have to hit, then the ring for the gold experiment has the same proportions as the one for copper, but it is enlarged by a factor equal to $Z_{\text{gold}}/Z_{\text{copper}}$. That is, not only is the radius of the ring greater by that factor, but unlike the rings on a normal dartboard, the thickness of the outer ring is also greater in proportion to its radius. When you take a geometric shape and scale it up in size like a photographic enlargement, its area is increased in proportion to the square of the enlargement factor, so the area of the dartboard ring in the gold experiment is greater by a factor equal to $(Z_{\text{gold}}/Z_{\text{copper}})^2$. Since the alphas are aimed entirely randomly, the chances of an alpha hitting the ring are in proportion to the area of the ring, which proves the equation given above.

As an example of the modern use of scattering experiments and cross-section measurements, you may have heard of the recent experimental evidence for the existence of a particle called the top quark. Of the twelve subatomic particles currently believed to be the smallest constituents of matter, six form a family called the quarks, distinguished from the other six by the intense attractive forces that make the quarks stick to each other. (The other six consist of the electron plus five other, more exotic particles.) The only two types of quarks found in naturally occurring matter are the "up quark" and "down quark," which are what protons and neutrons are made of, but four other types were theoretically predicted to exist, for a total of six. (The whimsical term "quark" comes from a line by James Joyce reading "Three quarks for master Mark.") Until recently, only five types of quarks had been proven to exist via experiments, and the sixth, the top quark, was only theorized. There was no hope of ever detecting a top quark directly, since it is radioactive, and only exists for a zillionth of a second before evaporating. Instead, the researchers searching for it at the Fermi National Accelerator Laboratory near Chicago measured cross-sections for scattering of nuclei off of other nuclei. The experiment was much like those of Rutherford and Chadwick, except that the incoming nuclei had to be boosted to much higher speeds in a particle accelerator. The resulting encounter with a target nucleus was so violent that both nuclei were completely demolished, but, as Einstein proved, energy can be converted into matter, and the energy of the collision creates a spray of exotic, radioactive particles, like the deadly shower of wood fragments produced by a cannon ball in an old naval battle. Among those particles were some top quarks. The cross-sections being measured were the cross-sections for the production of certain combinations of these secondary particles. However different the details, the principle was the same as that employed at the turn of the century: you smash things together and look at the fragments that fly off to see what was inside them. The approach has been compared to shooting a clock with a rifle and then studying the pieces that fly off to figure out how the clock worked.

### Discussion Questions

**A** The diagram, showing alpha particles being deflected by a gold nucleus, was drawn with the assumption that alpha particles came in on lines at many different distances from the nucleus. Why wouldn't they all come in along the same line, since they all came out through the same tube?

**B** Why does it make sense that, as shown in the figure, the trajectories that result in 19° and 20° scattering cross each other?

**C** Rutherford knew the velocity of the alpha particles emitted by radium, and guessed that the positively charged part of a gold atom had a charge of about +100$e$ (we now know it is +79$e$). Considering the fact that some alpha particles were deflected by 180°, how could he then use conservation of energy to derive an upper limit on the size of a gold nucleus? (For simplicity, assume the size of the alpha particle is negligible compared to that of the gold nucleus, and ignore the fact that the gold nucleus recoils a little from the collision, picking up a little kinetic energy.)

## 2.4 The structure of nuclei

### The proton

The fact that the nuclear charges were all integer multiples of $e$ suggested to many physicists that rather than being a pointlike object, the nucleus might contain smaller particles having individual charges of $+e$. Evidence in favor of this idea was not long in arriving. Rutherford reasoned that if he bombarded the atoms of a very light element with alpha particles, the small charge of the target nuclei would give a very weak repulsion. Perhaps those few alpha particles that happened to arrive on head-on collision courses would get so close that they would physically crash into some of the target nuclei. An alpha particle is itself a nucleus, so this would be a collision between two nuclei, and a violent one due to the high speeds involved. Rutherford hit pay dirt in an experiment with alpha particles striking a target containing nitrogen atoms. Charged particles were detected flying out of the target like parts flying off of cars in a high-speed crash. Measurements of the deflection of these particles in electric and magnetic fields showed that they had the same charge-to-mass ratio as singly-ionized hydrogen atoms. Rutherford concluded that these were the conjectured singly-charged particles that held the charge of the nucleus, and they were later named protons. The hydrogen nucleus consists of a single proton, and in general, an element's atomic number gives the number of protons contained in each of its nuclei. The mass of the proton is about 1800 times greater than the mass of the electron.

### The neutron

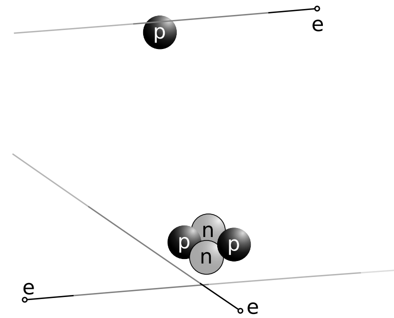
It would have been nice and simple if all the nuclei could have been built only from protons, but that couldn't be the case. If you spend a little time looking at a periodic table, you will soon notice that although some of the atomic masses are very nearly integer multiples of hydrogen's mass, many others are not. Even where the masses are close whole numbers, the masses of an element other than hydrogen is always greater than its atomic number, not equal to it. Helium, for instance, has two protons, but its mass is four times greater than that of hydrogen.

Chadwick cleared up the confusion by proving the existence of a new subatomic particle. Unlike the electron and proton, which are electrically charged, this particle is electrically neutral, and he named it the neutron. Chadwick's experiment has been described in detail in chapter 4 of book 2 of this series, but briefly the method was to expose a sample of the light element beryllium to a stream of alpha particles from a lump of radium. Beryllium has only four protons, so an alpha that happens to be aimed directly at a beryllium nucleus can actually hit it rather than being stopped short of a collision by electrical repulsion. Neutrons were observed as a new form of radiation emerging from the collisions, and Chadwick correctly inferred that they were previously unsuspected components of the nucleus that had been knocked out. As described in *Conservation Laws*, Chadwick also determined the mass of the neutron; it is very nearly the same as that of the proton.

To summarize, atoms are made of three types of particles:

| | *charge* | *mass in units of the proton's mass* | *location in atom* |
|---|---|---|---|
| proton | $+e$ | 1 | in nucleus |
| neutron | 0 | 1.001 | in nucleus |
| electron | $-e$ | 1/1836 | orbiting nucleus |

The existence of neutrons explained the mysterious masses of the elements. Helium, for instance, has a mass very close to four times greater than that of hydrogen. This is because it contains two neutrons in addition to its two protons. The mass of an atom is essentially determined by the total number of neutrons and protons. The total number of neutrons plus protons is therefore referred to as the atom's *mass number*.

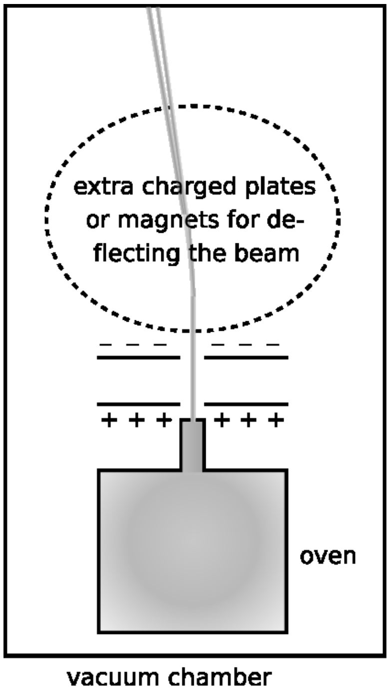



n / Examples of the construction of atoms: hydrogen (top) and helium (bottom). On this scale, the electrons' orbits would be the size of a college campus.

### Isotopes

We now have a clear interpretation of the fact that helium is close to four times more massive than hydrogen, and similarly for all the atomic masses that are close to an integer multiple of the mass of hydrogen. But what about copper, for instance, which had an atomic mass 63.5 times that of hydrogen? It didn't seem reasonable to think that it possessed an extra half of a neutron! The solution was found by measuring the mass-to-charge ratios of singly-ionized atoms (atoms with one electron removed). The technique is essentially that same as the one used by Thomson for cathode rays, except that whole atoms do not spontaneously leap out of the surface of an object as electrons sometimes do. Figure o shows an example of how the ions can be created and injected between the charged plates for acceleration.

Injecting a stream of copper ions into the device, we find a surprise — the beam splits into two parts! Chemists had elevated to dogma the assumption that all the atoms of a given element were identical, but we find that 69% of copper atoms have one mass, and 31% have another. Not only that, but both masses are very nearly integer multiples of the mass of hydrogen (63 and 65, respectively). Copper gets its chemical identity from the number of protons in its nucleus, 29, since chemical reactions work by electric forces. But apparently some copper atoms have $63 - 29 = 34$ neutrons while others have $65 - 29 = 36$. The atomic mass of copper, 63.5, reflects the proportions of the mixture of the mass-63 and mass-65 varieties. The different mass varieties of a given element are called $isotopes of$ that element.




o / A version of the Thomson apparatus modified for measuring the mass-to-charge ratios of ions rather than electrons. A small sample of the element in question, copper in our example, is boiled in the oven to create a thin vapor. (A vacuum pump is continuously sucking on the main chamber to keep it from accumulating enough gas to stop the beam of ions.) Some of the atoms of the vapor are ionized by a spark or by ultraviolet light. Ions that wander out of the nozzle and into the region between the charged plates are then accelerated toward the top of the figure. As in the Thomson experiment, mass-to-charge ratios are inferred from the deflection of the beam.

Isotopes can be named by giving the mass number as a subscript to the left of the chemical symbol, e.g., $^{65}$Cu. Examples:

| | *protons* | *neutrons* | *mass number* |
|---|---|---|---|
| $^{1}$H | 1 | 0 | 0+1 = 1 |
| $^{4}$He | 2 | 2 | 2+2 = 4 |
| $^{12}$C | 6 | 6 | 6+6 = 12 |
| $^{14}$C | 6 | 8 | 6+8 = 14 |
| $^{262}$Ha | 105 | 157 | 105+157 = 262 |

*self-check A*

Why are the positive and negative charges of the accelerating plates reversed in the isotope-separating apparatus compared to the Thomson apparatus? ▷ Answer, p. 195

Chemical reactions are all about the exchange and sharing of electrons: the nuclei have to sit out this dance because the forces of electrical repulsion prevent them from ever getting close enough to make contact with each other. Although the protons do have a vitally important effect on chemical processes because of their electrical forces, the neutrons can have no effect on the atom's chemical reactions. It is not possible, for instance, to separate $^{63}$Cu from $^{65}$Cu by chemical reactions. This is why chemists had never realized that different isotopes existed. (To be perfectly accurate, different isotopes do behave slightly differently because the more massive atoms move more sluggishly and therefore react with a tiny bit less intensity. This tiny difference is used, for instance, to separate out the isotopes of uranium needed to build a nuclear bomb. The smallness of this effect makes the separation process a slow and difficult one, which is what we have to thank for the fact that nuclear weapons have not been built by every terrorist cabal on the planet.)

## Sizes and shapes of nuclei

Matter is nearly all nuclei if you count by weight, but in terms of volume nuclei don't amount to much. The radius of an individual neutron or proton is very close to 1 fm (1 fm=$10^{-15}$ m), so even a big lead nucleus with a mass number of 208 still has a diameter of only about 13 fm, which is ten thousand times smaller than the diameter of a typical atom. Contrary to the usual imagery of the nucleus as a small sphere, it turns out that many nuclei are somewhat elongated, like an American football, and a few have exotic asymmetric shapes like pears or kiwi fruits.

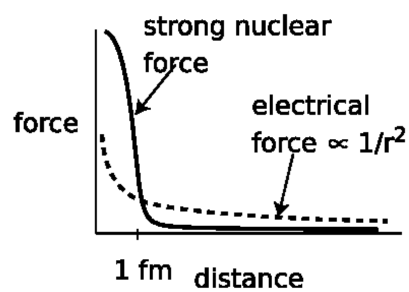
## Discussion Questions

**A** Suppose the entire universe was in a (very large) cereal box, and the nutritional labeling was supposed to tell a godlike consumer what percentage of the contents was nuclei. Roughly what would the percentage be like if the labeling was according to mass? What if it was by volume?

p / A nuclear power plant at Cattenom, France. Unlike the coal and oil plants that supply most of the U.S.'s electrical power, a nuclear power plant like this one releases no pollution or greenhouse gases into the Earth's atmosphere, and therefore doesn't contribute to global warming. The white stuff puffing out of this plant is non-radioactive water vapor. Although nuclear power plants generate long-lived nuclear waste, this waste arguably poses much less of a threat to the biosphere than greenhouse gases would.

## 2.5 The strong nuclear force, alpha decay and fission

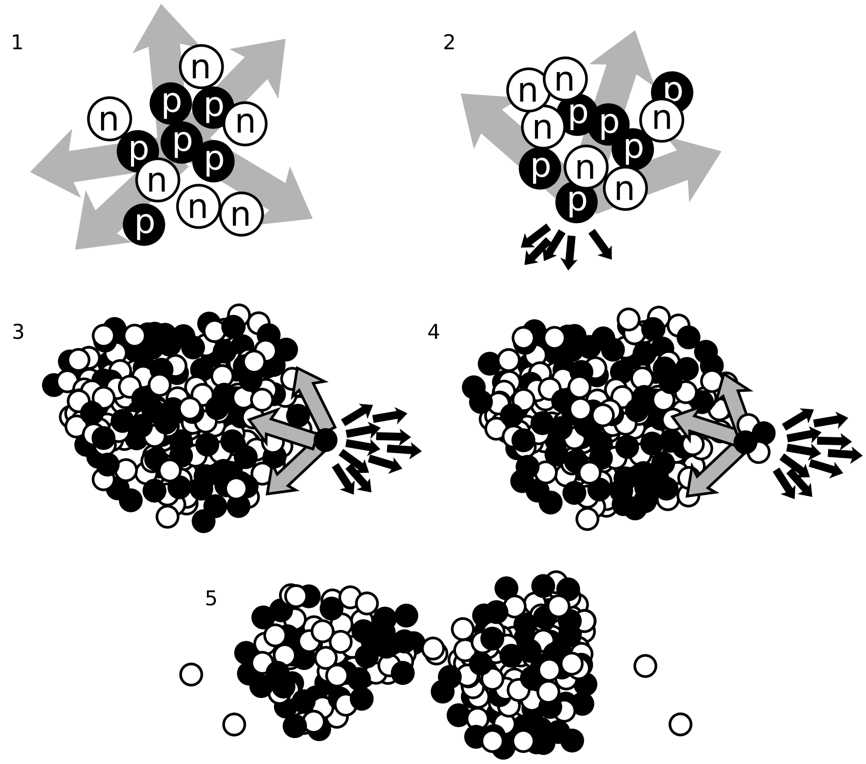
Once physicists realized that nuclei consisted of positively charged protons and uncharged neutrons, they had a problem on their hands. The electrical forces among the protons are all repulsive, so the nucleus should simply fly apart! The reason all the nuclei in your body are not spontaneously exploding at this moment is that there is another force acting. This force, called the $strong nuclear force$, is always attractive, and acts between neutrons and neutrons, neutrons and protons, and protons and protons with roughly equal strength. The strong nuclear force does not have any effect on electrons, which is why it does not influence chemical reactions.




q / The strong nuclear force cuts off very sharply at a range of about 1 fm.

Unlike electric forces, whose strengths are given by the simple Coulomb force law, there is no simple formula for how the strong nuclear force depends on distance. Roughly speaking, it is effective over ranges of $\sim$ 1 fm, but falls off extremely quickly at larger distances (much faster than $1/r^2$). Since the radius of a neutron or proton is about 1 fm, that means that when a bunch of neutrons and protons are packed together to form a nucleus, the strong nuclear force is effective only between neighbors.

Figure r illustrates how the strong nuclear force acts to keep ordinary nuclei together, but is not able to keep very heavy nuclei from breaking apart. In r/1, a proton in the middle of a carbon




r / 1. The forces cancel. 2. The forces don't cancel. 3. In a heavy nucleus, the large number of electrical repulsions can add up to a force that is comparable to the strong nuclear attraction. 4. Alpha emission. 5. Fission.

nucleus feels an attractive strong nuclear force (arrows) from each of its nearest neighbors. The forces are all in different directions, and tend to cancel out. The same is true for the repulsive electrical forces (not shown). In figure r/2, a proton at the edge of the nucleus has neighbors only on one side, and therefore all the strong nuclear forces acting on it are tending to pull it back in. Although all the electrical forces from the other five protons (dark arrows) are all pushing it out of the nucleus, they are not sufficient to overcome the strong nuclear forces.

In a very heavy nucleus, r/3, a proton that finds itself near the edge has only a few neighbors close enough to attract it significantly via the strong nuclear force, but every other proton in the nucleus exerts a repulsive electrical force on it. If the nucleus is large enough, the total electrical repulsion may be sufficient to overcome the attraction of the strong force, and the nucleus may spit out a proton. Proton emission is fairly rare, however; a more common type of radioactive decay[1] in heavy nuclei is alpha decay, shown in r/4. The imbalance of the forces is similar, but the chunk that is ejected is an

[1]Alpha decay is more common because an alpha particle happens to be a very stable arrangement of protons and neutrons.

alpha particle (two protons and two neutrons) rather than a single proton.

It is also possible for the nucleus to split into two pieces of roughly equal size, r/5, a process known as fission. Note that in addition to the two large fragments, there is a spray of individual neutrons. In a nuclear fission bomb or a nuclear fission reactor, some of these neutrons fly off and hit other nuclei, causing them to undergo fission as well. The result is a chain reaction.

When a nucleus is able to undergo one of these processes, it is said to be radioactive, and to undergo radioactive decay. Some of the naturally occurring nuclei on earth are radioactive. The term "radioactive" comes from Becquerel's image of rays radiating out from something, not from radio waves, which are a whole different phenomenon. The term "decay" can also be a little misleading, since it implies that the nucleus turns to dust or simply disappears – actually it is splitting into two new nuclei with an the same total number of neutrons and protons, so the term "radioactive transformation" would have been more appropriate. Although the original atom's electrons are mere spectators in the process of weak radioactive decay, we often speak loosely of "radioactive atoms" rather than "radioactive nuclei."

### Randomness in physics

How does an atom decide when to decay? We might imagine that it is like a termite-infested house that gets weaker and weaker, until finally it reaches the day on which it is destined to fall apart. Experiments, however, have not succeeded in detecting such "ticking clock" hidden below the surface; the evidence is that all atoms of a given isotope are absolutely identical. Why, then, would one uranium atom decay today while another lives for another million years? The answer appears to be that it is entirely random. We can make general statements about the average time required for a certain isotope to decay, or how long it will take for half the atoms in a sample to decay (its half-life), but we can never predict the behavior of a particular atom.

This is the first example we have encountered of an inescapable randomness in the laws of physics. If this kind of randomness makes you uneasy, you're in good company. Einstein's famous quote is "...I am convinced that He [God] does not play dice." Einstein's distaste for randomness, and his association of determinism with divinity, goes back to the Enlightenment conception of the universe as a gigantic piece of clockwork that only had to be set in motion initially by the Builder. Physics had to be entirely rebuilt in the 20th century to incorporate the fundamental randomness of physics, and this modern revolution is the topic of book 6 in this series. In particular, we will delay the mathematical development of the half-life concept until then.

## 2.6 The weak nuclear force; beta decay

All the nuclear processes we've discussed so far have involved rearrangements of neutrons and protons, with no change in the total number of neutrons or the total number of protons. Now consider the proportions of neutrons and protons in your body and in the planet earth: neutrons and protons are roughly equally numerous in your body's carbon and oxygen nuclei, and also in the nickel and iron that make up most of the earth. The proportions are about 50-50. But, as discussed in more detail in optional section 2.10, the only chemical elements produced in any significant quantities by the big bang[2] were hydrogen (about 90%) and helium (about 10%). If the early universe was almost nothing but hydrogen atoms, whose nuclei are protons, where did all those neutrons come from?

The answer is that there is another nuclear force, the weak nuclear force, that is capable of transforming neutrons into protons and vice-versa. Two possible reactions are

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu} \qquad \text{[electron decay]}$$

and

$$\mathrm{p} \rightarrow \mathrm{n} + \mathrm{e}^+ + \nu \qquad . \qquad \text{[positron decay]}$$

(There is also a third type called electron capture, in which a proton grabs one of the atom's electrons and they produce a neutron and a neutrino.)

Whereas alpha decay and fission are just a redivision of the previously existing particles, these reactions involve the destruction of one particle and the creation of three new particles that did not exist before.

There are three new particles here that you have never previously encountered. The symbol $\mathrm{e}^+$ stands for an antielectron, which is a particle just like the electron in every way, except that its electric charge is positive rather than negative. Antielectrons are also known as positrons. Nobody knows why electrons are so common in the universe and antielectrons are scarce. When an antielectron encounters an electron, they annihilate each other, producing gamma rays, and this is the fate of all the antielectrons that are produced by natural radioactivity on earth. Antielectrons are an example of antimatter. A complete atom of antimatter would consist of antiprotons, antielectrons, and antineutrons. Although individual particles of antimatter occur commonly in nature due to natural radioactivity and cosmic rays, only a few complete atoms of antihydrogen have ever been produced artificially.

The notation $\nu$ stands for a particle called a neutrino, and $\bar{\nu}$ means an antineutrino. Neutrinos and antineutrinos have no electric charge (hence the name).

[2]The evidence for the big bang theory of the origin of the universe was discussed in book 3 of this series.

We can now list all four of the known fundamental forces of physics:

- gravity
- electromagnetism
- strong nuclear force
- weak nuclear force

The other forces we have learned about, such as friction and the normal force, all arise from electromagnetic interactions between atoms, and therefore are not considered to be fundamental forces of physics.

*Decay of* $^{212}$Pb *example 2*

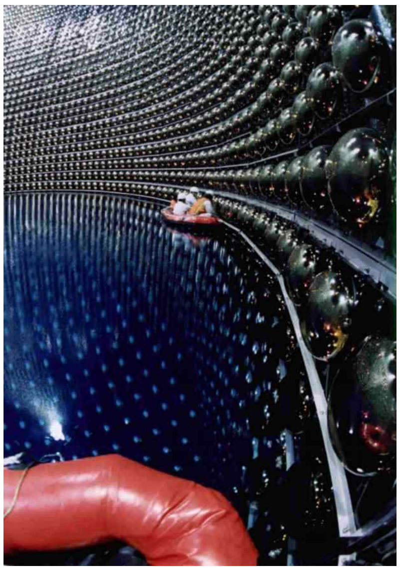
As an example, consider the radioactive isotope of lead $^{212}$Pb. It contains 82 protons and 130 neutrons. It decays by the process $n \rightarrow p + e^{-} + \bar{\nu}$ . The newly created proton is held inside the nucleus by the strong nuclear force, so the new nucleus contains 83 protons and 129 neutrons. Having 83 protons makes it the element bismuth, so it will be an atom of $^{212}$Bi.

In a reaction like this one, the electron flies off at high speed (typically close to the speed of light), and the escaping electrons are the things that make large amounts of this type of radioactivity dangerous. The outgoing electron was the first thing that tipped off scientists in the early 1900s to the existence of this type of radioactivity. Since they didn't know that the outgoing particles were electrons, they called them beta particles, and this type of radioactive decay was therefore known as beta decay. A clearer but less common terminology is to call the two processes electron decay and positron decay.

The neutrino or antineutrino emitted in such a reaction pretty much ignores all matter, because its lack of charge makes it immune to electrical forces, and it also remains aloof from strong nuclear interactions. Even if it happens to fly off going straight down, it is almost certain to make it through the entire earth without interacting with any atoms in any way. It ends up flying through outer space forever. The neutrino's behavior makes it exceedingly difficult to detect, and when beta decay was first discovered nobody realized that neutrinos even existed. We now know that the neutrino carries off some of the energy produced in the reaction, but at the time it seemed that the total energy afterwards (not counting the unsuspected neutrino's energy) was greater than the total energy before the reaction, violating conservation of energy. Physicists were getting ready to throw conservation of energy out the window as a basic law of physics when indirect evidence led them to the conclusion that neutrinos existed.

### The solar neutrino problem

What about these neutrinos? Why haven't you heard of them before? It's not because they're rare — a billion neutrinos pass through your body every microsecond, but until recently almost nothing was known about them. Produced as a side-effect of the nuclear reactions that power our sun and other stars, these ghostlike bits of matter are believed to be the most numerous particles in the universe. But they interact so weakly with ordinary matter that nearly all the neutrinos that enter the earth on one side will emerge from the other side of our planet without even slowing down.



s / This neutrino detector is in the process of being filled with ultrapure water.

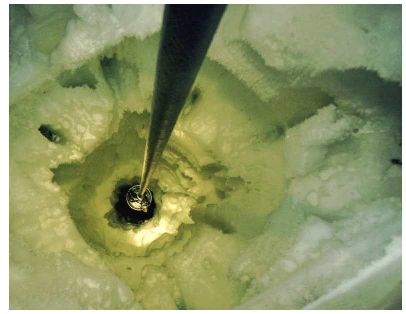
Our first real peek at the properties of the elusive neutrino has come from a huge detector in a played-out Japanese zinc mine, s. An international team of physicists outfitted the mineshaft with wall-to-wall light sensors, and then filled the whole thing with water so pure that you can see through it for a hundred meters, compared to only a few meters for typical tap water. Neutrinos stream through the 50 million liters of water continually, just as they flood everything else around us, and the vast majority never interact with a water molecule. A very small percentage, however, do annihilate themselves in the water, and the tiny flashes of light they produce can be detected by the beachball-sized vacuum tubes that line the darkened mineshaft. Most of the neutrinos around us come from the sun, but for technical reasons this type of water-based detector is more sensitive to the less common but more energetic neutrinos produced when cosmic ray particles strike the earth's atmosphere.

Neutrinos were already known to come in three "flavors," which can be distinguished from each other by the particles created when they collide with matter. An "electron-flavored neutrino" creates an ordinary electron when it is annihilated, while the two other types create more exotic particles called mu and tau particles. Think of the three types of neutrinos as chocolate, vanilla, and strawberry. When you buy a chocolate ice cream cone, you expect that it will keep being chocolate as you eat it. The unexpected finding from the Japanese experiment is that some of the neutrinos are changing flavor between the time when they are produced by a cosmic ray and the moment when they wink out of existence in the water. It's as though your chocolate ice cream cone transformed itself magically into strawberry while your back was turned.

How did the physicists figure out the change in flavor? The experiment detects some neutrinos originating in the atmosphere above Japan, and also many neutrinos coming from distant parts of the earth. A neutrino created above the Atlantic Ocean arrives in Japan from underneath, and the experiment can distinguish these upward-traveling neutrinos from the downward-moving local variety. They found that the mixture of neutrinos coming from below was different from the mixture arriving from above, with some of the electron-flavored and tau-flavored neutrinos having apparently changed into mu-flavored neutrinos during their voyage through the earth. The ones coming from above didn't have time to change flavors on their much shorter journey.

This is interpreted as evidence that the neutrinos are constantly changing back and forth among the three flavors. On theoretical grounds, it is believed that such a vibration can only occur if neutrinos have mass. Only a rough estimate of the mass is possible at this point: it appears that neutrinos have a mass somewhere in the neighborhood of one billionth of the mass of an electron, or about $10^{-39}$ kg.

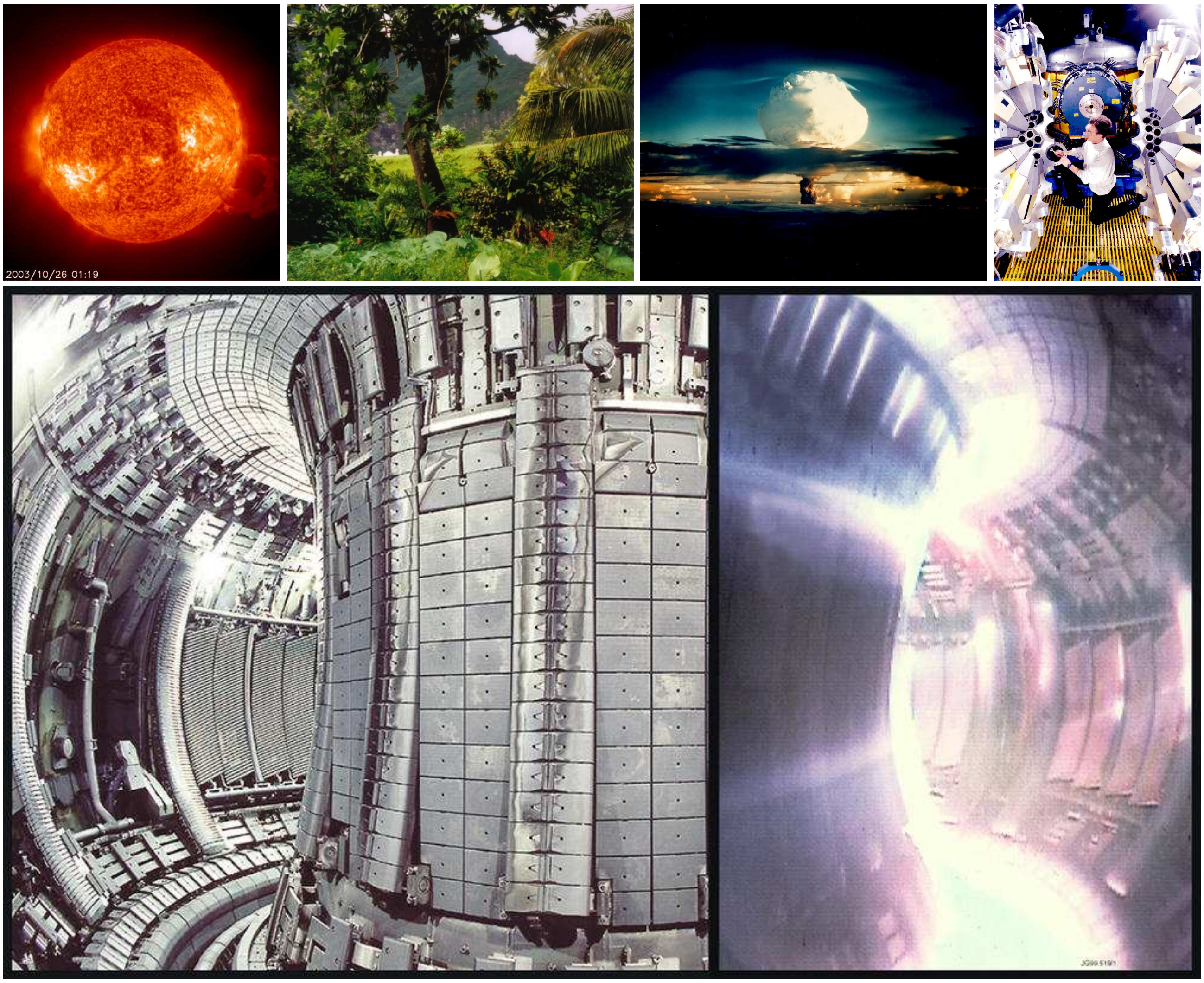
If the neutrino's mass is so tiny, does it even matter? It matters to astronomers. Neutrinos are the only particles that can be used to probe certain phenomena. For example, they are the only direct probes we have for testing our models of the core of our own sun, which is the source of energy for all life on earth. Once astronomers have a good handle on the basic properties of the neutrino, they can start thinking seriously about using them for astronomy. As of 2006, the mass of the neutrino has been confirmed by an accelerator-based experiment, and neutrino observatories have been operating for a few years in Antarctica, using huge volumes of natural ice in the same way that the water was used in the Japanese experiment.



t / A detector being lowered down a shaft at the IceCube neutrino telescope in Antarctica.

**A** In the reactions $n \rightarrow p + e^- + \bar{\nu}$ and $p \rightarrow n + e^+ + \nu$, verify that charge is conserved. In beta decay, when one of these reactions happens to a neutron or proton within a nucleus, one or more gamma rays may also be emitted. Does this affect conservation of charge? Would it be possible for some extra electrons to be released without violating charge conservation?

**B** When an antielectron and an electron annihilate each other, they produce two gamma rays. Is charge conserved in this reaction?




u / 1. Our sun's source of energy is nuclear fusion, so nuclear fusion is also the source of power for all life on earth, including, 2, this rain forest in Fatu-Hiva. 3. The first release of energy by nuclear fusion through human technology was the 1952 Ivy Mike test at the Enewetak Atoll. 4. This array of gamma-ray detectors is called GAMMASPHERE. During operation, the array is closed up, and a beam of ions produced by a particle accelerator strikes a target at its center, producing nuclear fusion reactions. The gamma rays can be studied for information about the structure of the fused nuclei, which are typically varieties not found in nature. 5. Nuclear fusion promises to be a clean, inexhaustible source of energy. However, the goal of commercially viable nuclear fusion power has remained elusive, due to the engineering difficulties involved in magnetically containing a plasma (ionized gas) at a sufficiently high temperature and density. This photo shows the experimental JET reactor, with the device opened up on the left, and in action on the right.

## 2.7 Fusion

As we have seen, heavy nuclei tend to fly apart because each proton is being repelled by every other proton in the nucleus, but is only attracted by its nearest neighbors. The nucleus splits up into two parts, and as soon as those two parts are more than about 1 fm apart, the strong nuclear force no longer causes the two fragments to attract each other. The electrical repulsion then accelerates them, causing them to gain a large amount of kinetic energy. This release of kinetic energy is what powers nuclear reactors and fission bombs.

It might seem, then, that the lightest nuclei would be the most stable, but that is not the case. Let's compare an extremely light nucleus like $^{4}$He with a somewhat heavier one, $^{16}$O. A neutron or proton in $^{4}$He can be attracted by the three others, but in $^{16}$O, it might have five or six neighbors attracting it. The $^{16}$O nucleus is therefore more stable.

It turns out that the most stable nuclei of all are those around nickel and iron, having about 30 protons and 30 neutrons. Just as a nucleus that is too heavy to be stable can release energy by splitting apart into pieces that are closer to the most stable size, light nuclei can release energy if you stick them together to make bigger nuclei that are closer to the most stable size. Fusing one nucleus with another is called nuclear fusion. Nuclear fusion is what powers our sun and other stars.

## 2.8 Nuclear energy and binding energies

In the same way that chemical reactions can be classified as exothermic (releasing energy) or endothermic (requiring energy to react), so nuclear reactions may either release or use up energy. The energies involved in nuclear reactions are greater by a huge factor. Thousands of tons of coal would have to be burned to produce as much energy as would be produced in a nuclear power plant by one kg of fuel.

Although nuclear reactions that use up energy (endothermic reactions) can be initiated in accelerators, where one nucleus is rammed into another at high speed, they do not occur in nature, not even in the sun. The amount of kinetic energy required is simply not available.

To find the amount of energy consumed or released in a nuclear reaction, you need to know how much nuclear interaction energy, $U_{nuc}$, was stored or released. Experimentalists have determined the amount of nuclear energy stored in the nucleus of every stable element, as well as many unstable elements. This is the amount of mechanical work that would be required to pull the nucleus apart into its individual neutrons and protons, and is known as the nuclear binding energy.

*A reaction occurring in the sun* *example 3*

The sun produces its energy through a series of nuclear fusion reactions. One of the reactions is

$$^{1}\text{H} + ^{2}\text{H} \rightarrow ^{3}\text{He} + \gamma$$

The excess energy is almost all carried off by the gamma ray (not by the kinetic energy of the helium-3 atom). The binding energies in units of pJ (picojoules) are:

| | |
|---|---|
| $^{1}$H | 0 J |
| $^{2}$H | 0.35593 pJ |
| $^{3}$He | 1.23489 pJ |

The total initial nuclear energy is 0 pJ+0.35593 pJ, and the final nuclear energy is 1.23489 pJ, so by conservation of energy, the gamma ray must carry off 0.87896 pJ of energy. The gamma ray is then absorbed by the sun and converted to heat.

*self-check B*

Why is the binding energy of $^{1}$H exactly equal to zero? ▷ Answer, p. 195

**Conversion of mass to energy and energy to mass**

If you add up the masses of the three particles produced in the reaction $n \rightarrow p + e^{-} + \bar{\nu}$, you will find that they do not equal the mass of the neutron, so mass is not conserved. An even more blatant example is the annihilation of an electron with a positron, $e^{-} + e^{+} \rightarrow 2\gamma$, in which the original mass is completely destroyed, since gamma rays have no mass. Nonconservation of mass is not just a property of nuclear reactions. It also occurs in chemical reactions, but the change in mass is too small to detect with ordinary laboratory balances.

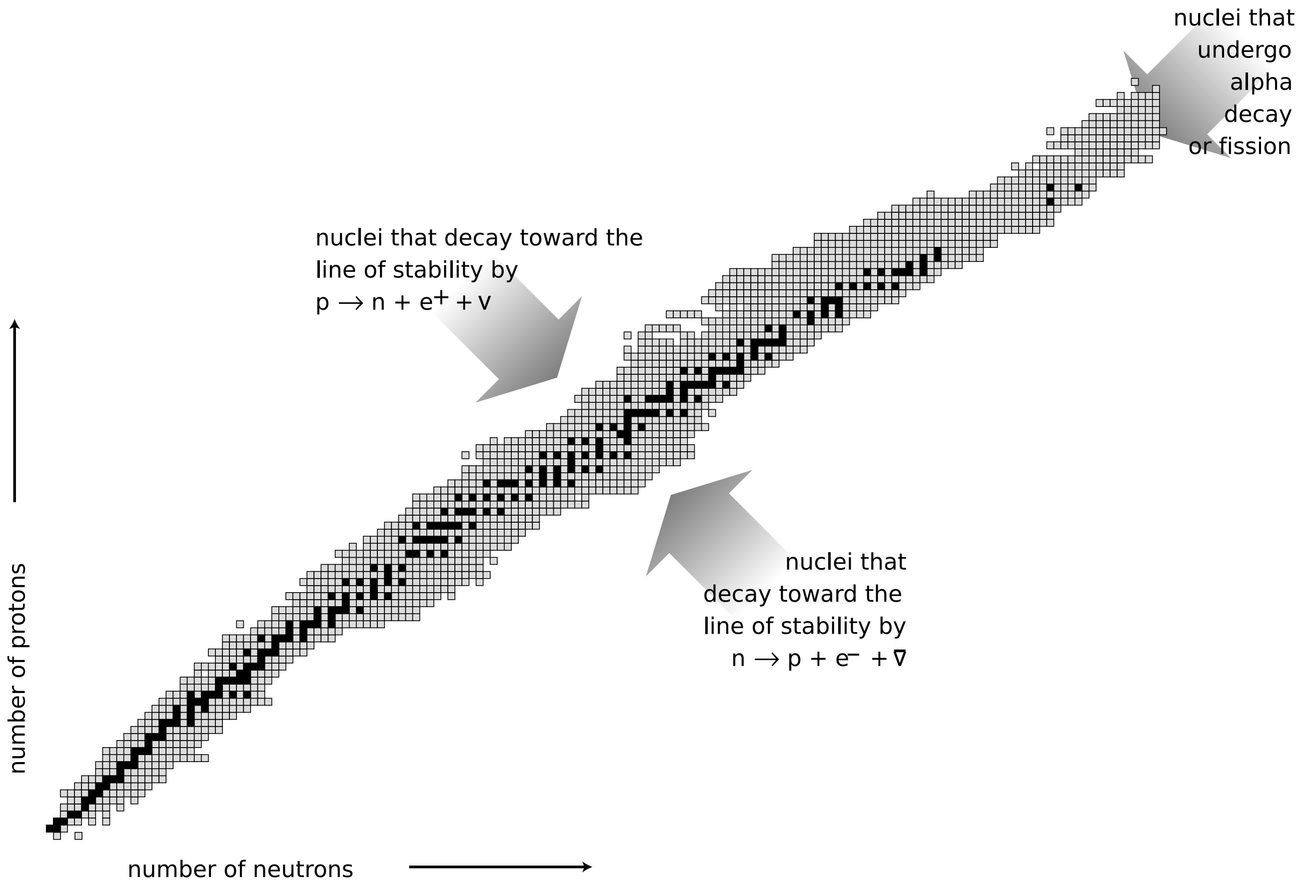
The reason why mass is not being conserved is that mass is being converted to energy, according to Einstein's celebrated equation $E = mc^2$, in which $c$ stands for the speed of light. In the reaction $e^- + e^+ \rightarrow 2\gamma$, for instance, imagine for simplicity that the electron and positron are moving very slowly when they collide, so there is no significant amount of energy to start with. We are starting with mass and no energy, and ending up with two gamma rays that possess energy but no mass. Einstein's $E = mc^2$ tells us that the conversion factor between mass and energy is equal to the square of the speed of light. Since $c$ is a big number, the amount of energy consumed or released by a chemical reaction only shows up as a tiny change in mass. But in nuclear reactions, which involve large amounts of energy, the change in mass may amount to as much as one part per thousand. Note that in this context, $c$ is not necessarily the speed of any of the particles. We are just using its numerical value as a conversion factor. Note also that $E = mc^2$ does not mean that an object of mass $m$ has a kinetic energy equal to $mc^2$; the energy being described by $E = mc^2$ is the energy you could release if you destroyed the particle and converted its mass entirely into energy, and that energy would be in addition to any kinetic or potential energy the particle had.

Have we now been cheated out of two perfectly good conservation laws, the laws of conservation of mass and of energy? No, it's just that according to Einstein, the conserved quantity is $E + mc^2$, not $E$ or $m$ individually. The quantity $E + mc^2$ is referred to as the mass-energy, and no violation of the law of conservation of mass-energy has yet been observed. In most practical situations, it is a perfectly reasonable to treat mass and energy as separately conserved quantities.

It is now easy to explain why isolated protons (hydrogen nuclei) are found in nature, but neutrons are only encountered in the interior of a nucleus, not by themselves. In the process $n \rightarrow p + e^- + \bar{\nu}$, the total final mass is less than the mass of the neutron, so mass is being converted into energy. In the beta decay of a proton, $p \rightarrow n + e^+ + \nu$, , the final mass is greater than the initial mass, so some energy needs to be supplied for conversion into mass. A proton sitting by itself in a hydrogen atom cannot decay, since it has no source of energy. Only protons sitting inside nuclei can decay, and only then if the difference in potential energy between the original nucleus and the new nucleus would result in a release of energy. But any isolated neutron that is created in natural or artificial reactions will decay within a matter of seconds, releasing some energy.

The equation $E = mc^2$ occurs naturally as part of Einstein's theory of special relativity, which is not what we are studying right now. This brief treatment is only meant to clear up the issue of where the mass was going in some of the nuclear reactions we were discussing.

Figure v is a compact way of showing the vast variety of the nuclei. Each box represents a particular number of neutrons and protons. The black boxes are nuclei that are stable, i.e., that would require an input of energy in order to change into another. The gray boxes show all the unstable nuclei that have been studied experimentally. Some of these last for billions of years on the aver-



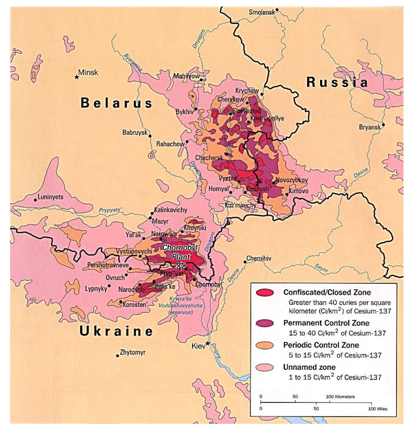

v / The known nuclei, represented on a chart of proton number versus neutron number. Note the two nuclei in the bottom row with zero protons. One is simply a single neutron. The other is a cluster of four neutrons. This "tetraneutron" was reported, unexpectedly, to be a bound system in results from a 2002 experiment. The result is controversial. If correct, it implies the existence of a heretofore unsuspected type of matter, the neutron droplet, which we can think of as an atom with no protons or electrons.

age before decaying and are found in nature, but most have much shorter average lifetimes, and can only be created and studied in the laboratory.

The curve along which the stable nuclei lie is called the line of stability. Nuclei along this line have the most stable proportion of neutrons to protons. For light nuclei the most stable mixture is about 50-50, but we can see that stable heavy nuclei have two or three times more neutrons than protons. This is because the electrical repulsions of all the protons in a heavy nucleus add up to a powerful force that would tend to tear it apart. The presence of a large number of neutrons increases the distances among the protons, and also increases the number of attractions due to the strong nuclear force.

## 2.9 Biological effects of ionizing radiation

w / An abandoned village near Chernobyl.





x / A map showing levels of radiation near the site of the Chernobyl disaster. At the boundary of the most highly contaminated (bright red) areas, people would be exposed to about 1300 millirem per year, or about four times the natural background level. In the pink areas, which are still densely populated, the exposure is comparable to the natural level found in a high-altitude city such as Denver.

As a science educator, I find it frustrating that nowhere in the massive amount of journalism devoted to the Chernobyl disaster does one ever find any numerical statements about the amount of radiation to which people have been exposed. Anyone mentally capable of understanding sports statistics or weather reports ought to be able to understand such measurements, as long as something like the following explanatory text was inserted somewhere in the article:

Radiation exposure is measured in units of millirems. The average person is exposed to about 200 millirems each year from natural background sources.

With this context, people would be able to come to informed conclusions based on statements such as, "Children in Finland received an average dose of __________ millirems above natural background levels because of the Chernobyl disaster."

A millirem, or mrem, is, of course, a thousandth of a rem, but what is a rem? It measures the amount of energy per kilogram deposited in the body by ionizing radiation, multiplied by a "quality factor" to account for the different health hazards posed by alphas, betas, gammas, neutrons, and other types of radiation. Only ionizing radiation is counted, since nonionizing radiation simply heats one's body rather than killing cells or altering DNA. For instance, alpha particles are typically moving so fast that their kinetic energy is sufficient to ionize thousands of atoms, but it is possible for an alpha particle to be moving so slowly that it would not have enough kinetic energy to ionize even one atom.

Notwithstanding the pop culture images of the Incredible Hulk and Godzilla, it is not possible for a multicellular animal to become "mutated" as a whole. In most cases, a particle of ionizing radiation will not even hit the DNA, and even if it does, it will only affect the DNA of a single cell, not every cell in the animal's body. Typically, that cell is simply killed, because the DNA becomes unable to function properly. Once in a while, however, the DNA may be altered so as to make that cell cancerous. For instance, skin cancer can be caused by UV light hitting a single skin cell in the body of a sunbather. If that cell becomes cancerous and begins reproducing uncontrollably, she will end up with a tumor twenty years later.

Other than cancer, the only other dramatic effect that can result from altering a single cell's DNA is if that cell happens to be a sperm or ovum, which can result in nonviable or mutated offspring. Men are relatively immune to reproductive harm from radiation, because their sperm cells are replaced frequently. Women are more vulnerable because they keep the same set of ova as long as they live.

A whole-body exposure of 500,000 mrem will kill a person within a week or so. Luckily, only a small number of humans have ever been exposed to such levels: one scientist working on the Manhattan Project, some victims of the Nagasaki and Hiroshima explosions, and 31 workers at Chernobyl. Death occurs by massive killing of cells, especially in the blood-producing cells of the bone marrow.

Lower levels, on the order of 100,000 mrem, were inflicted on some people at Nagasaki and Hiroshima. No acute symptoms result from this level of exposure, but certain types of cancer are significantly more common among these people. It was originally expected that the radiation would cause many mutations resulting in birth defects, but very few such inherited effects have been observed.

A great deal of time has been spent debating the effects of very low levels of ionizing radiation. A medical x-ray, for instance, may result in a dose on the order of a 100 mrem above background, i.e., less than a doubling of the normal background level. Similar doses in excess of the average background level may be received by people living at high altitudes or people with high concentrations of radon gas in their houses. Unfortunately (or fortunately, depending on how you look at it), the added risks of cancer or birth defects resulting from these levels of exposure are extremely small, and therefore nearly impossible to measure. As with many suspected carcinogenic chemicals, the only practical method of estimating risks is to give laboratory animals doses many orders of magnitude greater, and then assume that the health risk is directly proportional to the dose. Under these assumptions, the added risk posed by a dental x-ray or radon in one's basement is negligible on a personal level, and is only significant in terms of a slight increase in cancer throughout the population. As a matter of social policy, excess radiation exposure is not a significant public health problem compared to car accidents or tobacco smoking.

### Discussion Questions

**A** Should the quality factor for neutrinos be very small, because they mostly don't interact with your body?

**B** Would an alpha source be likely to cause different types of cancer depending on whether the source was external to the body or swallowed in contaminated food? What about a gamma source?

## 2.10 ⋆ The creation of the elements

### Creation of hydrogen and helium in the Big Bang

Did all the chemical elements we're made of come into being in the big bang?[3] Temperatures in the first microseconds after the big bang were so high that atoms and nuclei could not hold together at all. After things had cooled down enough for nuclei and atoms to exist, there was a period of about three minutes during which the temperature and density were high enough for fusion to occur, but not so high that atoms could hold together. We have a good, detailed understanding of the laws of physics that apply under these conditions, so theorists are able to say with confidence that the only element heavier than hydrogen that was created in significant quantities was helium.

### We are stardust

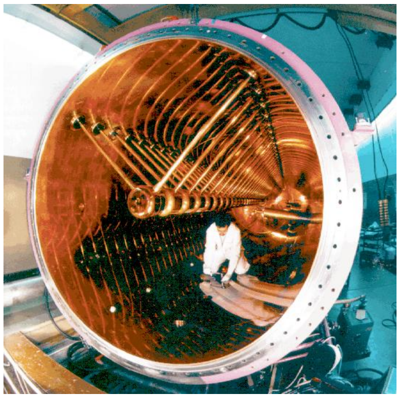
In that case, where did all the other elements come from? Astronomers came up with the answer. By studying the combinations of wavelengths of light, called spectra, emitted by various stars, they had been able to determine what kinds of atoms they contained. (We will have more to say about spectra at the end of this book.) They found that the stars fell into two groups. One type was nearly 100% hydrogen and helium, while the other contained 99% hydrogen and helium and 1% other elements. They interpreted these as two generations of stars. The first generation had formed out of clouds of gas that came fresh from the big bang, and their composition reflected that of the early universe. The nuclear fusion reactions by which they shine have mainly just increased the proportion of helium relative to hydrogen, without making any heavier elements.

y / The Crab Nebula is a remnant of a supernova explosion. Almost all the elements our planet is made of originated in such explosions.

The members of the first generation that we see today, however, are only those that lived a long time. Small stars are more miserly with their fuel than large stars, which have short lives. The large stars of the first generation have already finished their lives. Near the end of its lifetime, a star runs out of hydrogen fuel and undergoes a series of violent and spectacular reorganizations as it fuses heavier and heavier elements. Very large stars finish this sequence of events by undergoing supernova explosions, in which some of their material is flung off into the void while the rest collapses into an exotic object such as a black hole or neutron star.

The second generation of stars, of which our own sun is an example, condensed out of clouds of gas that had been enriched in heavy elements due to supernova explosions. It is those heavy elements that make up our planet and our bodies.

[3] The evidence for the big bang theory of the origin of the universe was discussed in book 3 of this series.



z / Construction of the UNILAC accelerator in Germany, one of whose uses is for experiments to create very heavy artificial elements. In such an experiment, fusion products recoil through a device called SHIP (not shown) that separates them based on their charge-to-mass ratios — it is essentially just a scaled-up version of Thomson's apparatus. A typical experiment runs for several months, and out of the billions of fusion reactions induced during this time, only one or two may result in the production of superheavy atoms. In all the rest, the fused nucleus breaks up immediately. SHIP is used to identify the small number of "good" reactions and separate them from this intense background.

### Artificial synthesis of heavy elements

Elements up to uranium, atomic number 92, were created by these astronomical processes. Beyond that, the increasing electrical repulsion of the protons leads to shorter and shorter half-lives. Even if a supernova a billion years ago did create some quantity of an element such as Berkelium, number 97, there would be none left in the Earth's crust today. The heaviest elements have all been created by artificial fusion reactions in accelerators. As of 2006, the heaviest element that has been created is 116.[4]

Although the creation of a new element, i.e., an atom with a novel number of protons, has historically been considered a glamorous accomplishment, to the nuclear physicist the creation of an atom with a hitherto unobserved number of neutrons is equally important. The greatest neutron number reached so far is 179. One tantalizing goal of this type of research is the theoretical prediction that there might be an island of stability beyond the previously explored tip of the chart of the nuclei shown in section 2.8. Just as certain numbers of electrons lead to the chemical stability of the noble gases (helium, neon, argon, ...), certain numbers of neutrons and protons lead to a particularly stable packing of orbits. Calculations dating back to the 1960's have hinted that there might be relatively stable nuclei having approximately 114 protons and 184 neutrons. The isotopes of elements 114 and 116 that have been produced so far have had half-lives in the second or millosecond range. This may not seem like very long, but lifetimes in the microsecond range are more typical for the superheavy elements that have previously been discovered. There is even speculation that certain superheavy isotopes would be stable enough to be produced in quantities that could for instance be weighed and used in chemical reactions.

[4] An earlier claim of the creation of element 116 by a group at Berkeley turned out to be a case of scientific fraud, but the element was later produced by a different group, at Dubna, Russia.

## Summary

### Selected Vocabulary

| | |
|---|---|
| alpha particle | a form of radioactivity consisting of helium nuclei |
| beta particle | a form of radioactivity consisting of electrons |
| gamma ray | a form of radioactivity consisting of a very high-frequency form of light |
| proton | a positively charged particle, one of the types that nuclei are made of |
| neutron | an uncharged particle, the other types that nuclei are made of |
| isotope | one of the possible varieties of atoms of a given element, having a certain number of neutrons |
| atomic number | the number of protons in an atom's nucleus; determines what element it is |
| atomic mass | the mass of an atom |
| mass number | the number of protons plus the number of neutrons in a nucleus; approximately proportional to its atomic mass |
| strong nuclear force | the force that holds nuclei together against electrical repulsion |
| weak nuclear force | the force responsible for beta decay |
| beta decay | the radioactive decay of a nucleus via the reaction $n \rightarrow p + e^- + \bar{\nu}$ or $p \rightarrow n + e^+ + \nu$; so called because an electron or antielectron is also known as a beta particle |
| alpha decay | the radioactive decay of a nucleus via emission of an alpha particle |
| fission | the radioactive decay of a nucleus by splitting into two parts |
| fusion | a nuclear reaction in which two nuclei stick together to form one bigger nucleus |
| millirem | a unit for measuring a person's exposure to radioactivity |

### Notation

| | |
|---|---|
| $e^-$ | an electron |
| $e^+$ | an antielectron; just like an electron, but with positive charge |
| n | a neutron |
| p | a proton |
| $\nu$ | a neutrino |
| $\bar{\nu}$ | a neutrino |

### Other Terminology and Notation

| | |
|---|---|
| $Z$ . . . . . . . . . | atomic number (number of protons in a nucleus) |
| $N$ . . . . . . . . . | number of neutrons in a nucleus |
| $A$ . . . . . . . . . | mass number ($N + Z$) |

### Summary

Rutherford and Marsden observed that some alpha particles from a beam striking a thin gold foil came back at angles up to 180 degrees. This could not be explained in the then-favored raisin -cookie model of the atom, and led to the adoption of the planetary model of the atom, in which the electrons orbit a tiny, positively-charged nucleus. Further experiments showed that the nucleus itself was a cluster of positively-charged protons and uncharged neutrons.

Radioactive nuclei are those that can release energy. The most common types of radioactivity are alpha decay (the emission of a helium nucleus), beta decay (the transformation of a neutron into a proton or vice-versa), and gamma decay (the emission of a type of very-high-frequency light). Stars are powered by nuclear fusion reactions, in which two light nuclei collide and form a bigger nucleus, with the release of energy.

Human exposure to ionizing radiation is measured in units of millirem. The typical person is exposed to about 200 mrem worth of natural background radiation per year.

### Exploring Further

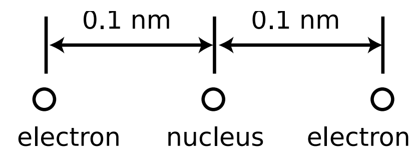
**The First Three Minutes**, Steven Weinberg. This book describes the first three minutes of the universe's existence.

## Problems

### Key

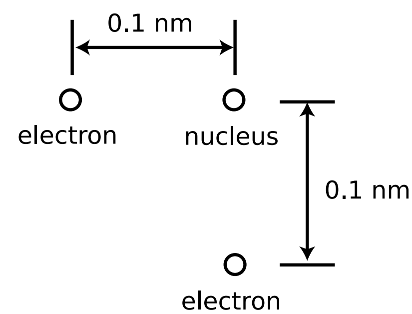
- √ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ⋆ A difficult problem.

**1** A helium atom finds itself momentarily in this arrangement. Find the direction and magnitude of the force acting on the right-hand electron. The two protons in the nucleus are so close together ($\sim 1$ fm) that you can consider them as being right on top of each other. √




Problem 1.

**2** The helium atom of problem 1 has some new experiences, goes through some life changes, and later on finds itself in the configuration shown here. What are the direction and magnitude of the force acting on the bottom electron? (Draw a sketch to make clear the definition you are using for the angle that gives direction.) √




Problem 2.

**3** Suppose you are holding your hands in front of you, 10 cm apart.
(a) Estimate the total number of electrons in each hand. √
(b) Estimate the total repulsive force of all the electrons in one hand on all the electrons in the other. √
(c) Why don't you feel your hands repelling each other?
(d) Estimate how much the charge of a proton could differ in magnitude from the charge of an electron without creating a noticeable force between your hands.

**4** Suppose that a proton in a lead nucleus wanders out to the surface of the nucleus, and experiences a strong nuclear force of about 8 kN from the nearby neutrons and protons pulling it back in. Compare this numerically to the repulsive electrical force from the other protons, and verify that the net force is attractive. A lead nucleus is very nearly spherical, and is about 6.5 fm in radius. √

**5** The subatomic particles called muons behave exactly like electrons, except that a muon's mass is greater by a factor of 206.77. Muons are continually bombarding the Earth as part of the stream of particles from space known as cosmic rays. When a muon strikes an atom, it can displace one of its electrons. If the atom happens to be a hydrogen atom, then the muon takes up an orbit that is on the average 206.77 times closer to the proton than the orbit of the ejected electron. How many times greater is the electric force experienced by the muon than that previously felt by the electron? √

**6** The nuclear process of beta decay by electron capture is described parenthetically in section 2.6. The reaction is $\text{p}+\text{e}^- \rightarrow \text{n}+\nu$.
(a) Show that charge is conserved in this reaction.
(b) Conversion between energy and mass is discussed in the optional topic on page 66. Based on these ideas, explain why electron capture doesn't occur in hydrogen atoms. (If it did, matter wouldn't

exist!)

▷ Solution, p. 196

**7** $^{234}$Pu decays either by electron decay or by alpha decay. (A given $^{234}$Pu nucleus may do either one; it's random.) What are the isotopes created as products fo these two modes of decay?

# Chapter 3
# Circuits, Part 1

Madam, what good is a baby? *Michael Faraday, when asked by Queen Victoria what the electrical devices in his lab were good for*

A few years ago, my wife and I bought a house with Character, Character being a survival mechanism that houses have evolved in order to convince humans to agree to much larger mortgage payments than they'd originally envisioned. Anyway, one of the features that gives our house Character is that it possesses, built into the wall of the family room, a set of three pachinko machines. These are Japanese gambling devices sort of like vertical pinball machines. (The legal papers we got from the sellers hastened to tell us that they were "for amusement purposes only.") Unfortunately, only one of the three machines was working when we moved in, and it soon died on us. Having become a pachinko addict, I decided to fix it, but that was easier said than done. The inside is a veritable Rube Goldberg mechanism of levers, hooks, springs, and chutes. My hormonal pride, combined with my Ph.D. in physics, made me certain of success, and rendered my eventual utter failure all the more demoralizing.

Contemplating my defeat, I realized how few complex mechanical devices I used from day to day. Apart from our cars and my saxophone, every technological tool in our modern life-support system was electronic rather than mechanical.

## 3.1 Current

### Unity of all types of electricity

We are surrounded by things we have been *told* are "electrical," but it's far from obvious what they have in common to justify being grouped together. What relationship is there between the way socks cling together and the way a battery lights a lightbulb? We have been told that both an electric eel and our own brains are somehow electrical in nature, but what do they have in common?

British physicist Michael Faraday (1791-1867) set out to address this problem. He investigated electricity from a variety of sources — including electric eels! — to see whether they could all produce the same effects, such as shocks and sparks, attraction and repulsion. "Heating" refers, for example, to the way a lightbulb filament gets hot enough to glow and emit light. Magnetic induction is an effect discovered by Faraday himself that connects electricity and magnetism. We will not study this effect, which is the basis for the electric generator, in detail until later in the book.

| source | effect | | | |
|---|---|---|---|---|
| | shocks | sparks | attraction and repulsion | heating |
| rubbing | √ | √ | √ | √ |
| battery | √ | √ | √ | √ |
| animal | √ | √ | (√) | √ |
| magnetically induced | √ | √ | √ | √ |

a / *Gymnotus carapo*, a knifefish, uses electrical signals to sense its environment and to communicate with others of its species.

The table shows a summary of some of Faraday's results. Check marks indicate that Faraday or his close contemporaries were able to verify that a particular source of electricity was capable of producing a certain effect. (They evidently failed to demonstrate attraction and repulsion between objects charged by electric eels, although modern workers have studied these species in detail and been able to understand all their electrical characteristics on the same footing as other forms of electricity.)

Faraday's results indicate that there is nothing fundamentally different about the types of electricity supplied by the various sources. They are all able to produce a wide variety of identical effects. Wrote Faraday, "The general conclusion which must be drawn from this collection of facts is that electricity, whatever may be its source, is identical in its nature."

If the types of electricity are the same thing, what thing is that? The answer is provided by the fact that all the sources of electricity can cause objects to repel or attract each other. We use the word "charge" to describe the property of an object that allows it to participate in such electrical forces, and we have learned that charge is present in matter in the form of nuclei and electrons. Evidently all these electrical phenomena boil down to the motion of charged particles in matter.

### Electric current

If the fundamental phenomenon is the motion of charged particles, then how can we define a useful numerical measurement of it? We might describe the flow of a river simply by the velocity of the water, but velocity will not be appropriate for electrical purposes because we need to take into account how much charge the moving particles have, and in any case there are no practical devices sold at Radio Shack that can tell us the velocity of charged particles. Experiments show that the intensity of various electrical effects is related to a different quantity: the number of coulombs of charge that pass by a certain point per second. By analogy with the flow of water, this quantity is called the electric *current*, $I$. Its units of coulombs/second are more conveniently abbreviated as amperes, 1 A=1 C/s. (In informal speech, one usually says "amps.")

b / André Marie Ampère (1775-1836).

The main subtlety involved in this definition is how to account for the two types of charge. The stream of water coming from a hose is made of atoms containing charged particles, but it produces none of the effects we associate with electric currents. For example, you do not get an electrical shock when you are sprayed by a hose. This type of experiment shows that the effect created by the motion of one type of charged particle can be canceled out by the motion of the opposite type of charge in the same direction. In water, every oxygen atom with a charge of $+8e$ is surrounded by eight electrons with charges of $-e$, and likewise for the hydrogen atoms.

We therefore refine our definition of current as follows:

**definition of electric current**
When charged particles are exchanged between regions of space A and B, the electric current flowing from A to B is

$$I = \frac{\Delta q}{\Delta t} \qquad ,$$

where $\Delta q$ is the change in region B's total charge occurring over a period of time $\Delta t$.

In the garden hose example, your body picks up equal amounts of positive and negative charge, resulting in no change in your total charge, so the electrical current flowing into you is zero.

*Interpretation of $\Delta q/\Delta t$* *example 1*

▷ How should the expression $\Delta q/\Delta t$ be interpreted when the current isn't constant?

▷ You've seen lots of equations of this form before: $v = \Delta x/\Delta t$, $F = \Delta p/\Delta t$, etc. These are all descriptions of rates of change, and they all require that the rate of change be constant. If the rate of change isn't constant, you instead have to use the slope of the tangent line on a graph. The slope of a tangent line is equivalent to a derivative in calculus; applications of calculus are discussed in section 3.6.

*Ions moving across a cell membrane* *example 2*

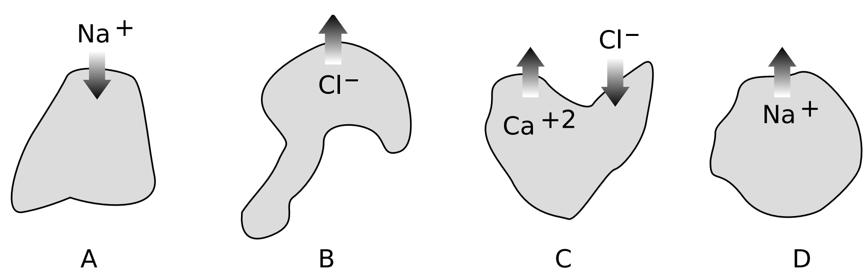
▷ Figure c shows ions, labeled with their charges, moving in or out through the membranes of three cells. If the ions all cross the membranes during the same interval of time, how would the currents into the cells compare with each other?

▷ Cell A has positive current going into it because its charge is increased, i.e., has a positive value of $\Delta q$.

Cell B has the same current as cell A, because by losing one unit of negative charge it also ends up increasing its own total charge by one unit.

Cell C's total charge is reduced by three units, so it has a large negative current going into it.

Cell D loses one unit of charge, so it has a small negative current into it.




c / Example 2

It may seem strange to say that a negatively charged particle going one way creates a current going the other way, but this is quite ordinary. As we will see, currents flow through metal wires via the motion of electrons, which are negatively charged, so the direction of motion of the electrons in a circuit is always opposite to the direction of the current. Of course it would have been convenient of Benjamin Franklin had defined the positive and negative signs of charge the opposite way, since so many electrical devices are based on metal wires.

*Number of electrons flowing through a lightbulb* *example 3*

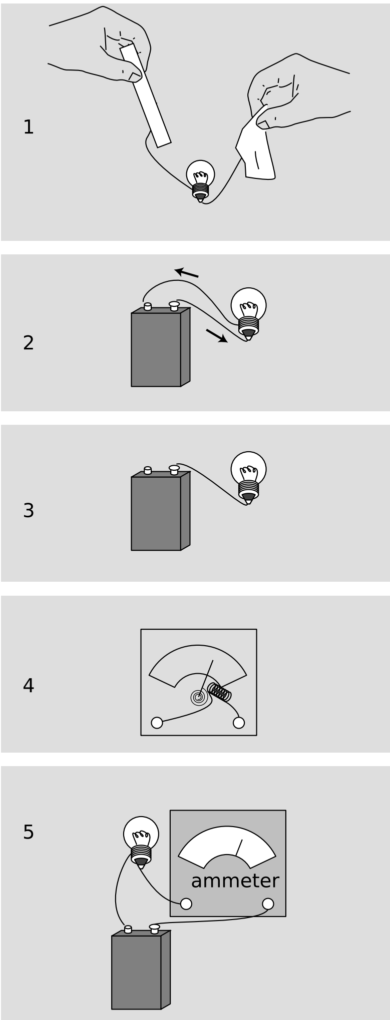
▷ If a lightbulb has 1.0 A flowing through it, how many electrons will pass through the filament in 1.0 s?

▷ We are only calculating the number of electrons that flow, so we can ignore the positive and negative signs. Solving for $\Delta q = I\Delta t$ gives a charge of 1.0 C flowing in this time interval. The number of electrons is

$$\begin{aligned}\text{number of electrons} &= \text{coulombs} \times \frac{\text{electrons}}{\text{coulomb}} \\ &= \text{coulombs} \Big/ \frac{\text{coulombs}}{\text{electron}} \\ &= 1.0\ \text{C}/e \\ &= 6.2 \times 10^{18}\end{aligned}$$

## 3.2 Circuits

How can we put electric currents to work? The only method of controlling electric charge we have studied so far is to charge different substances, e.g., rubber and fur, by rubbing them against each other. Figure d/1 shows an attempt to use this technique to light a lightbulb. This method is unsatisfactory. True, current will flow through the bulb, since electrons can move through metal wires, and the excess electrons on the rubber rod will therefore come through the wires and bulb due to the attraction of the positively charged fur and the repulsion of the other electrons. The problem is that after a zillionth of a second of current, the rod and fur will both have run out of charge. No more current will flow, and the lightbulb will go out.




d / 1. Static electricity runs out quickly. 2. A practical circuit. 3. An open circuit. 4. How an ammeter works. 5. Measuring the current with an ammeter.

Figure d/2 shows a setup that works. The battery pushes charge through the circuit, and recycles it over and over again. (We will have more to say later in this chapter about how batteries work.) This is called a *complete circuit*. Today, the electrical use of the word "circuit" is the only one that springs to mind for most people, but the original meaning was to travel around and make a round trip, as when a circuit court judge would ride around the boondocks, dispensing justice in each town on a certain date.

Note that an example like d/3 does not work. The wire will quickly begin acquiring a net charge, because it has no way to get rid of the charge flowing into it. The repulsion of this charge will make it more and more difficult to send any more charge in, and soon the electrical forces exerted by the battery will be canceled out completely. The whole process would be over so quickly that the filament would not even have enough time to get hot and glow. This is known as an *open circuit*. Exactly the same thing would happen if the complete circuit of figure d/2 was cut somewhere with a pair of scissors, and in fact that is essentially how an ordinary light switch works: by opening up a gap in the circuit.

The definition of electric current we have developed has the great virtue that it is easy to measure. In practical electrical work, one almost always measures current, not charge. The instrument used to measure current is called an *ammeter*. A simplified ammeter, d/4, simply consists of a coiled-wire magnet whose force twists an iron needle against the resistance of a spring. The greater the current, the greater the force. Although the construction of ammeters may differ, their use is always the same. We break into the path of the electric current and interpose the meter like a tollbooth on a road, d/5. There is still a complete circuit, and as far as the battery and bulb are concerned, the ammeter is just another segment of wire.

Does it matter where in the circuit we place the ammeter? Could we, for instance, have put it in the left side of the circuit instead of the right? Conservation of charge tells us that this can make no difference. Charge is not destroyed or "used up" by the lightbulb, so we will get the same current reading on either side of it. What is "used up" is energy stored in the battery, which is being converted into heat and light energy.

## 3.3 Voltage

### The volt unit

Electrical circuits can be used for sending signals, storing information, or doing calculations, but their most common purpose by far is to manipulate energy, as in the battery-and-bulb example of the previous section. We know that lightbulbs are rated in units of watts, i.e., how many joules per second of energy they can convert into heat and light, but how would this relate to the flow of charge as measured in amperes? By way of analogy, suppose your friend, who didn't take physics, can't find any job better than pitching bales of hay. The number of calories he burns per hour will certainly depend on how many bales he pitches per minute, but it will also be proportional to how much mechanical work he has to do on each bale. If his job is to toss them up into a hayloft, he will got tired a lot more quickly than someone who merely tips bales off a loading dock into trucks. In metric units,

$$\frac{\text{joules}}{\text{second}} = \frac{\text{haybales}}{\text{second}} \times \frac{\text{joules}}{\text{haybale}} \qquad .$$

e / Alessandro Volta (1745-1827).

Similarly, the rate of energy transformation by a battery will not just depend on how many coulombs per second it pushes through a circuit but also on how much mechanical work it has to do on each coulomb of charge:

$$\frac{\text{joules}}{\text{second}} = \frac{\text{coulombs}}{\text{second}} \times \frac{\text{joules}}{\text{coulomb}}$$

or

$$\text{power} = \text{current} \times \text{work per unit charge} \qquad .$$

Units of joules per coulomb are abbreviated as *volts*, 1 V=1 J/C, named after the Italian physicist Alessandro Volta. Everyone knows that batteries are rated in units of volts, but the voltage concept is more general than that; it turns out that voltage is a property of every point in space. To gain more insight, let's think more carefully about what goes on in the battery and bulb circuit.

### The voltage concept in general

To do work on a charged particle, the battery apparently must be exerting forces on it. How does it do this? Well, the only thing that can exert an electrical force on a charged particle is another charged particle. It's as though the haybales were pushing and pulling each other into the hayloft! This is potentially a horribly complicated situation. Even if we knew how much excess positive or negative charge there was at every point in the circuit (which realistically we don't) we would have to calculate zillions of forces using Coulomb's law, perform all the vector additions, and finally calculate how much work was being done on the charges as they moved along. To make things even more scary, there is more than one type of charged particle that moves: electrons are what move in the wires and the bulb's filament, but ions are the moving charge carriers inside the battery. Luckily, there are two ways in which we can simplify things:

**The situation is unchanging.** Unlike the imaginary setup in which we attempted to light a bulb using a rubber rod and a piece of fur, this circuit maintains itself in a steady state (after perhaps a microsecond-long period of settling down after the circuit is first assembled). The current is steady, and as charge flows out of any area of the circuit it is replaced by the same amount of charge flowing in. The amount of excess positive or negative charge in any part of the circuit therefore stays constant. Similarly, when we watch a river flowing, the water goes by but the river doesn't disappear.

**Force depends only on position.** Since the charge distribution is not changing, the total electrical force on a charged particle depends only on its own charge and on its location. If another charged particle of the same type visits the same location later on, it will feel exactly the same force.

The second observation tells us that there is nothing all that different about the experience of one charged particle as compared to another's. If we single out one particle to pay attention to, and figure out the amount of work done on it by electrical forces as it goes from point A to point $B$ along a certain path, then this is the same amount of work that will be done on any other charged particles of the same type as it follows the same path. For the sake of visualization, let's think about the path that starts at one terminal of the battery, goes through the light bulb's filament, and ends at the other terminal. When an object experiences a force that depends only on its position (and when certain other, technical conditions are satisfied), we can define an electrical energy associated with the position of that object. The amount of work done on the particle by electrical forces as it moves from A to B equals the drop in electrical energy between A and B. This electrical energy is what is being converted into other forms of energy such as heat and light. We therefore define voltage in general as electrical energy per unit charge:

**definition of voltage difference**

The difference in voltage between two points in space is defined as

$$\Delta V = \Delta U_{elec}/q \qquad ,$$

where $\Delta U_{elec}$ is the change in the electrical energy of a particle with charge $q$ as it moves from the initial point to the final point.

The amount of power dissipated (i.e., rate at which energy is transformed by the flow of electricity) is then given by the equation

$$P = I\Delta V \qquad .$$

*Energy stored in a battery* — *example 4*

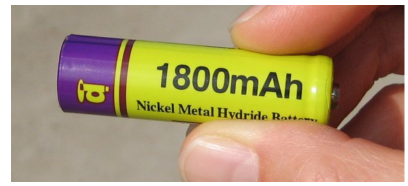
▷ The 1.2 V rechargeable battery in figure f is labeled 1800 milliamp-hours. What is the maximum amount of energy the battery can store?

▷ An ampere-hour is a unit of current multiplied by a unit of time. Current is charge per unit time, so an ampere-hour is in fact a funny unit of *charge*:

$$\begin{aligned}(1\text{ A})(1\text{ hour}) &= (1\text{ C/s})(3600\text{ s})\\ &= 3600\text{ C}\end{aligned}$$

1800 milliamp-hours is therefore $1800 \times 10^{-3} \times 3600\text{ C} = 6.5 \times 10^3\text{ C}$. That's a huge number of charged particles, but the total loss of electrical energy will just be their total charge multiplied by the voltage difference across which they move:

$$\begin{aligned}\Delta U_{elec} &= q\Delta V\\ &= (6.5 \times 10^3\text{ C})(1.2\text{ V})\\ &= 7.8\text{ kJ}\end{aligned}$$



f / Example 4.

*Units of volt-amps* — *example 5*

▷ Doorbells are often rated in volt-amps. What does this combination of units mean?

▷ Current times voltage gives units of power, $P = I\Delta V$, so volt-amps are really just a nonstandard way of writing watts. They are telling you how much power the doorbell requires.

*Power dissipated by a battery and bulb* — *example 6*

▷ If a 9.0-volt battery causes 1.0 A to flow through a lightbulb, how much power is dissipated?

▷ The voltage rating of a battery tells us what voltage difference $\Delta V$ it is designed to maintain between its terminals.

$$\begin{aligned}P &= I\,\Delta V\\ &= 9.0\text{ A}\cdot\text{V}\\ &= 9.0\,\frac{\text{C}}{\text{s}}\cdot\frac{\text{J}}{\text{C}}\\ &= 9.0\text{ J/s}\\ &= 9.0\text{ W}\end{aligned}$$

The only nontrivial thing in this problem was dealing with the units. One quickly gets used to translating common combinations like A · V into simpler terms.

Here are a few questions and answers about the voltage concept.

*Question:* OK, so what *is* voltage, really?
*Answer:* A device like a battery has positive and negative charges inside it that push other charges around the outside circuit. A higher-voltage battery has denser charges in it, which will do more work on each charged particle that moves through the outside circuit.

To use a gravitational analogy, we can put a paddlewheel at the bottom of either a tall waterfall or a short one, but a kg of water that falls through the greater gravitational energy difference will have more energy to give up to the paddlewheel at the bottom.

*Question:* Why do we define voltage as electrical energy divided by charge, instead of just defining it as electrical energy?
*Answer:* One answer is that it's the only definition that makes the equation $P = I\Delta V$ work. A more general answer is that we want to be able to define a voltage difference between any two points in space without having to know in advance how much charge the particles moving between them will have. If you put a nine-volt battery on your tongue, then the charged particles that move across your tongue and give you that tingly sensation are not electrons but ions, which may have charges of $+e$, $-2e$, or practically anything. The manufacturer probably expected the battery to be used mostly in circuits with metal wires, where the charged particles that flowed would be electrons with charges of $-e$. If the ones flowing across your tongue happen to have charges of $-2e$, the electrical energy difference for them will be twice as much, but dividing by their charge of $-2e$ in the definition of voltage will still give a result of 9 $V$.

*Question:* Are there two separate roles for the charged particles in the circuit, a type that sits still and exerts the forces, and another that moves under the influence of those forces?
*Answer:* No. Every charged particle simultaneously plays both roles. Newton's third law says that any particle that has an electrical forces acting on it must also be exerting an electrical force back on the other particle. There are no "designated movers" or "designated force-makers."

*Question:* Why does the definition of voltage only refer to voltage *differences*?
*Answer:* It's perfectly OK to define voltage as $V = U_{elec}/q$. But recall that it is only *differences* in interaction energy, $U$, that have direct physical meaning in physics. Similarly, voltage differences are really more useful than absolute voltages. A voltmeter measures voltage differences, not absolute voltages.

## Discussion Questions

**A** A roller coaster is sort of like an electric circuit, but it uses gravitational forces on the cars instead of electric ones. What would a high-voltage roller coaster be like? What would a high-current roller coaster be like?

**B** Criticize the following statements:

> "He touched the wire, and 10000 volts went through him."
>
> "That battery has a charge of 9 volts."
>
> "You used up the charge of the battery."

**C** When you touch a 9-volt battery to your tongue, both positive and negative ions move through your saliva. Which ions go which way?

**D** I once touched a piece of physics apparatus that had been wired incorrectly, and got a several-thousand-volt voltage difference across my hand. I was not injured. For what possible reason would the shock have had insufficient power to hurt me?

## 3.4 Resistance

### Resistance

So far we have simply presented it as an observed fact that a battery-and-bulb circuit quickly settles down to a steady flow, but why should it? Newton's second law, $a = F/m$, would seem to predict that the steady forces on the charged particles should make them whip around the circuit faster and faster. The answer is that as charged particles move through matter, there are always forces, analogous to frictional forces, that resist the motion. These forces need to be included in Newton's second law, which is really $a = F_{total}/m$, not $a = F/m$. If, by analogy, you push a crate across the floor at constant speed, i.e., with zero acceleration, the total force on it must be zero. After you get the crate going, the floor's frictional force is exactly canceling out your force. The chemical energy stored in your body is being transformed into heat in the crate and the floor, and no longer into an increase in the crate's kinetic energy. Similarly, the battery's internal chemical energy is converted into heat, not into perpetually increasing the charged particles' kinetic energy. Changing energy into heat may be a nuisance in some circuits, such as a computer chip, but it is vital in a lightbulb, which must get hot enough to glow. Whether we like it or not, this kind of heating effect is going to occur any time charged particles move through matter.

What determines the amount of heating? One flashlight bulb designed to work with a 9-volt battery might be labeled 1.0 watts, another 5.0. How does this work? Even without knowing the details of this type of friction at the atomic level, we can relate the heat dissipation to the amount of current that flows via the equation $P = I\Delta \mathrm{V}$. If the two flashlight bulbs can have two different values of $P$ when used with a battery that maintains the same $\Delta V$, it must be that the 5.0-watt bulb allows five times more current to flow through it.

g / Georg Simon Ohm (1787-1854).

For many substances, including the tungsten from which lightbulb filaments are made, experiments show that the amount of current that will flow through it is directly proportional to the voltage difference placed across it. For an object made of such a substance, we define its electrical *resistance* as follows:

**definition of resistance**
If an object inserted in a circuit displays a current flow proportional to the voltage difference across it, then we define its resistance as the constant ratio

$$R = \Delta V/I$$

The units of resistance are volts/ampere, usually abbreviated as ohms, symbolized with the capital Greek letter omega, $\Omega$.

*Resistance of a lightbulb* *example 7*

▷ A flashlight bulb powered by a 9-volt battery has a resistance of 10 $\Omega$. How much current will it draw?

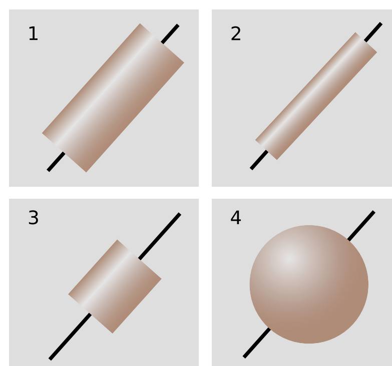
▷ Solving the definition of resistance for $I$, we find

$$\begin{aligned} I &= \Delta V/R \\ &= 0.9\ \text{V}/\Omega \\ &= 0.9\ \text{V}/(\text{V}/\text{A}) \\ &= 0.9\ \text{A} \end{aligned}$$

Ohm's law states that many substances, including many solids and some liquids, display this kind of behavior, at least for voltages that are not too large. The fact that Ohm's law is called a "law" should not be taken to mean that all materials obey it, or that it has the same fundamental importance as Newton's laws, for example. Materials are called *ohmic* or *nonohmic*, depending on whether they obey Ohm's law.

If objects of the same size and shape made from two different ohmic materials have different resistances, we can say that one material is more resistive than the other, or equivalently that it is less conductive. Materials, such as metals, that are very conductive are said to be good *conductors*. Those that are extremely poor conductors, for example wood or rubber, are classified as *insulators*. There is no sharp distinction between the two classes of materials. Some, such as silicon, lie midway between the two extremes, and are called semiconductors.

On an intuitive level, we can understand the idea of resistance by making the sounds "hhhhhh" and "ffffff." To make air flow out of your mouth, you use your diaphragm to compress the air in your chest. The pressure difference between your chest and the air outside your mouth is analogous to a voltage difference. When you make the "h" sound, you form your mouth and throat in a way that allows air to flow easily. The large flow of air is like a large current. Dividing by a large current in the definition of resistance means that we get a small resistance. We say that the small resistance of your mouth and throat allows a large current to flow. When you make the "f" sound, you increase the resistance and cause a smaller current to flow.




h / Four objects made of the same substance have different resistances.

Note that although the resistance of an object depends on the substance it is made of, we cannot speak simply of the "resistance of gold" or the "resistance of wood." Figure h shows four examples of objects that have had wires attached at the ends as electrical connections. If they were made of the same substance, they would all nevertheless have different resistances because of their different sizes and shapes. A more detailed discussion will be more natural in the context of the following chapter, but it should not be too surprising that the resistance of h/2 will be greater than that of h/1

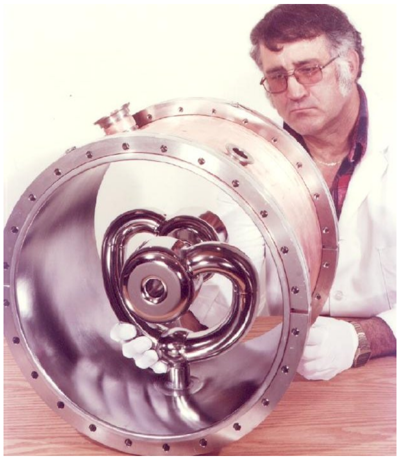
— the image of water flowing through a pipe, however incorrect, gives us the right intuition. Object h/3 will have a smaller resistance than h/1 because the charged particles have less of it to get through.

### Superconductors

All materials display some variation in resistance according to temperature (a fact that is used in thermostats to make a thermometer that can be easily interfaced to an electric circuit). More spectacularly, most metals have been found to exhibit a sudden change to *zero* resistance when cooled to a certain critical temperature. They are then said to be superconductors. Theoretically, superconductors should make a great many exciting devices possible, for example coiled-wire magnets that could be used to levitate trains. In practice, the critical temperatures of all metals are very low, and the resulting need for extreme refrigeration has made their use uneconomical except for such specialized applications as particle accelerators for physics research.

But scientists have recently made the surprising discovery that certain ceramics are superconductors at less extreme temperatures. The technological barrier is now in finding practical methods for making wire out of these brittle materials. Wall Street is currently investing billions of dollars in developing superconducting devices for cellular phone relay stations based on these materials. In 2001, the city of Copenhagen replaced a short section of its electrical power trunks with superconducing cables, and they are now in operation and supplying power to customers.

There is currently no satisfactory theory of superconductivity in general, although superconductivity in metals is understood fairly well. Unfortunately I have yet to find a fundamental explanation of superconductivity in metals that works at the introductory level.



i / A superconducting segment of the ATLAS accelerator at Argonne National Laboratory near Chicago. It is used to accelerate beams of ions to a few percent of the speed of light for nuclear physics research. The shiny silver-colored surfaces are made of the element niobium, which is a superconductor at relatively high temperatures compared to other metals — relatively high meaning the temperature of liquid helium! The beam of ions passes through the holes in the two small cylinders on the ends of the curved rods. Charge is shuffled back and forth between them at a frequency of 12 million cycles per second, so that they take turns being positive and negative. The positively charged beam consists of short spurts, each timed so that when it is in one of the segments it will be pulled forward by negative charge on the cylinder in front of it and pushed forward by the positively charged one behind. The huge currents involved (see example 9 on page 99) would quickly melt any metal that was not superconducting, but in a superconductor they produce no heat at all.

## Constant voltage throughout a conductor

The idea of a superconductor leads us to the question of how we should expect an object to behave if it is made of a very good conductor. Superconductors are an extreme case, but often a metal wire can be thought of as a perfect conductor, for example if the parts of the circuit other than the wire are made of much less conductive materials. What happens if $R$ equals zero in the equation $R = \Delta V/I$? The result of dividing two numbers can only be zero if the number on top equals zero. This tells us that if we pick any two points in a perfect conductor, the voltage difference between them must be zero. In other words, the entire conductor must be at the same voltage.

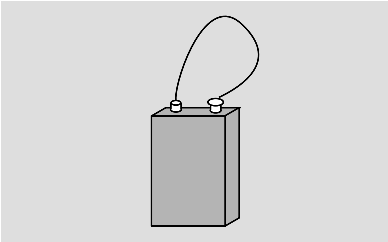
Constant voltage means that no work would be done on a charge as it moved from one point in the conductor to another. If zero work was done only along a certain path between two specific points, it might mean that positive work was done along part of the path and negative work along the rest, resulting in a cancellation. But there is no way that the work could come out to be zero for all possible paths unless the electrical force on a charge was in fact zero at every point. Suppose, for example, that you build up a static charge by scuffing your feet on a carpet, and then you deposit some of that charge onto a doorknob, which is a good conductor. How can all that charge be in the doorknob without creating any electrical force at any point inside it? The only possible answer is that the charge moves around until it has spread itself into just the right configuration so that the forces exerted by all the little bits of excess surface charge on any charged particle within the doorknob exactly canceled out.

We can explain this behavior if we assume that the charge placed on the doorknob eventually settles down into a stable equilibrium. Since the doorknob is a conductor, the charge is free to move through it. If it was free to move and any part of it did experience a nonzero total force from the rest of the charge, then it would move, and we would not have an equilibrium.

It also turns out that charge placed on a conductor, once it reaches its equilibrium configuration, is entirely on the surface, not on the interior. We will not prove this fact formally, but it is intuitively reasonable. Suppose, for instance, that the net charge on the conductor is negative, i.e., it has an excess of electrons. These electrons all repel each other, and this repulsion will tend to push them onto the surface, since being on the surface allows them to be as far apart as possible.

## Short circuits

So far we have been assuming a perfect conductor. What if it is a good conductor, but not a perfect one? Then we can solve for $\Delta V = IR$. An ordinary-sized current will make a very small result when we multiply it by the resistance of a good conductor such as a metal wire. The voltage throughout the wire will then be nearly constant. If, on the other hand, the current is extremely large, we can have a significant voltage difference. This is what happens in a *short-circuit:* a circuit in which a low-resistance pathway connects the two sides of a voltage source. Note that this is much more specific than the popular use of the term to indicate any electrical malfunction at all. If, for example, you short-circuit a 9-volt battery as shown in figure j, you will produce perhaps a thousand amperes of current, leading to a very large value of $P = I\Delta V$. The wire gets hot!



j / Short-circuiting a battery. Warning: you can burn yourself this way or start a fire! If you want to try this, try making the connection only very briefly, use a low-voltage battery, and avoid touching the battery or the wire, both of which will get hot.

*self-check A*

What would happen to the battery in this kind of short circuit? ▷ Answer, p. 195

## Resistors

Inside any electronic gadget you will see quite a few little circuit elements like the one shown below. These *resistors* are simply a cylinder of ohmic material with wires attached to the end.

k / The symbol used in schematics to represent a resistor.

At this stage, most students have a hard time understanding why resistors would be used inside a radio or a computer. We obviously want a lightbulb or an electric stove to have a circuit element that resists the flow of electricity and heats up, but heating is undesirable in radios and computers. Without going too far afield, let's use a mechanical analogy to get a general idea of why a resistor would be used in a radio.

The main parts of a radio receiver are an antenna, a tuner for selecting the frequency, and an amplifier to strengthen the signal sufficiently to drive a speaker. The tuner resonates at the selected frequency, just as in the examples of mechanical resonance discussed in book 3 of this series. The behavior of a mechanical resonator depends on three things: its inertia, its stiffness, and the amount of friction or damping. The first two parameters locate the peak of the resonance curve, while the damping determines the width of the resonance. In the radio tuner we have an electrically vibrating system that resonates at a particular frequency. Instead of a physical object moving back and forth, these vibrations consist of electrical currents that flow first in one direction and then in the other. In a mechanical system, damping means taking energy out of the vibration in the form of heat, and exactly the same idea applies to an electrical system: the resistor supplies the damping, and therefore controls the width of the resonance. If we set out to eliminate all resistance in the tuner circuit, by not building in a resistor and by somehow getting rid of all the inherent electrical resistance of the wires, we would have a useless radio. The tuner's resonance would be so narrow that we could never get close enough to the right frequency to bring in the station. The roles of inertia and stiffness are played by other circuit elements we have not discusses (a capacitor and a coil).

Many electrical devices are based on electrical resistance and Ohm's law, even if they do not have little components in them that look like the usual resistor. The following are some examples.

### Lightbulb

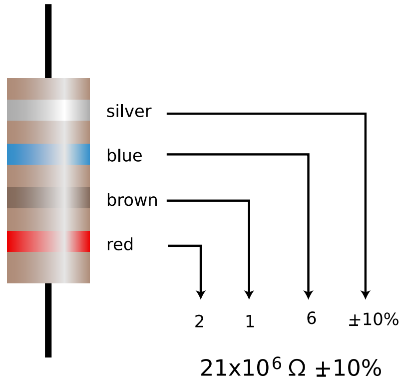
There is nothing special about a lightbulb filament — you can easily make a lightbulb by cutting a narrow waist into a metallic gum wrapper and connecting the wrapper across the terminals of a 9-volt battery. The trouble is that it will instantly burn out. Edison solved this technical challenge by encasing the filament in an evacuated bulb, which prevented burning, since burning requires oxygen.

### Polygraph

The polygraph, or "lie detector," is really just a set of meters for recording physical measures of the subject's psychological stress, such as sweating and quickened heartbeat. The real-time sweat measurement works on the principle that dry skin is a good insulator, but sweaty skin is a conductor. Of course a truthful subject may become nervous simply because of the situation, and a practiced liar may not even break a sweat. The method's practitioners claim that they can tell the difference, but you should think twice before allowing yourself to be polygraph tested. Most U.S. courts exclude all polygraph evidence, but some employers attempt to screen out dishonest employees by polygraph testing job applicants, an abuse that ranks with such pseudoscience as handwriting analysis.

### Fuse

A fuse is a device inserted in a circuit tollbooth-style in the same manner as an ammeter. It is simply a piece of wire made of metals having a relatively low melting point. If too much current passes through the fuse, it melts, opening the circuit. The purpose is to make sure that the building's wires do not carry so much current that they themselves will get hot enough to start a fire. Most modern houses use circuit breakers instead of fuses, although fuses are still common in cars and small devices. A circuit breaker is a switch operated by a coiled-wire magnet, which opens the circuit when enough current flows. The advantage is that once you turn off some




l / An example of a resistor with a color code.

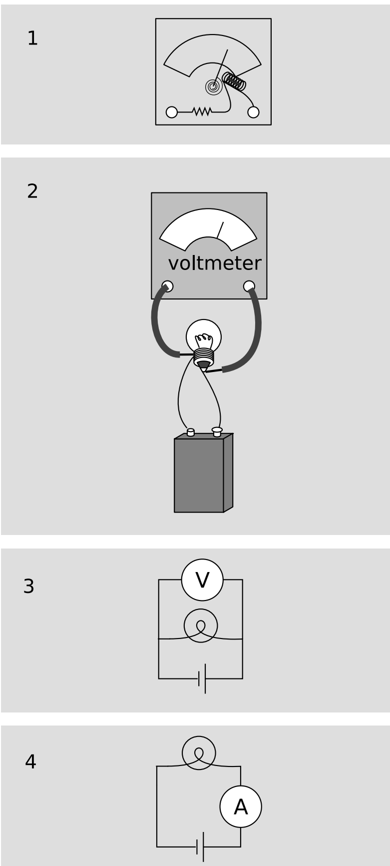
| color | meaning |
|---|---|
| black | 0 |
| brown | 1 |
| red | 2 |
| orange | 3 |
| yellow | 4 |
| green | 5 |
| blue | 6 |
| violet | 7 |
| gray | 8 |
| white | 9 |
| silver | ±10% |
| gold | ±5% |

m / Color codes used on resistors.

of the appliances that were sucking up too much current, you can immediately flip the switch closed. In the days of fuses, one might get caught without a replacement fuse, or even be tempted to stuff aluminum foil in as a replacement, defeating the safety feature.

### Voltmeter

A voltmeter is nothing more than an ammeter with an additional high-value resistor through which the current is also forced to flow. Ohm's law relates the current through the resistor is related directly to the voltage difference across it, so the meter can be calibrated in units of volts based on the known value of the resistor. The voltmeter's two probes are touched to the two locations in a circuit between which we wish to measure the voltage difference, n/2. Note how cumbersome this type of drawing is, and how difficult it can be to tell what is connected to what. This is why electrical drawing are usually shown in schematic form. Figure n/3 is a schematic representation of figure n/2.



n / 1. A simplified diagram of how a voltmeter works. 2. Measuring the voltage difference across a lightbulb. 3. The same setup drawn in schematic form. 4. The setup for measuring current is different.

The setups for measuring current and voltage are different. When we are measuring current, we are finding "how much stuff goes through," so we place the ammeter where all the current is forced to go through it. Voltage, however, is not "stuff that goes through," it is a measure of electrical energy. If an ammeter is like the meter that measures your water use, a voltmeter is like a measuring stick that tells you how high a waterfall is, so that you can determine how much energy will be released by each kilogram of falling water. We do not want to force the water to go through the measuring stick! The arrangement in figure n/3 is a *parallel* circuit: one in there are "forks in the road" where some of the current will flow one way and some will flow the other. Figure n/4 is said to be wired in *series:* all the current will visit all the circuit elements one after the other. We will deal with series and parallel circuits in more detail in the following chapter.

If you inserted a voltmeter incorrectly, in series with the bulb and battery, its large internal resistance would cut the current down so low that the bulb would go out. You would have severely disturbed the behavior of the circuit by trying to measure something about it.

Incorrectly placing an ammeter in parallel is likely to be even more disconcerting. The ammeter has nothing but wire inside it to provide resistance, so given the choice, most of the current will flow through it rather than through the bulb. So much current will flow through the ammeter, in fact, that there is a danger of burning out the battery or the meter or both! For this reason, most ammeters have fuses or circuit breakers inside. Some models will trip their circuit breakers and make an audible alarm in this situation, while others will simply blow a fuse and stop working until you replace it.

### Discussion Questions

**A** In figure n/1, would it make any difference in the voltage measurement if we touched the voltmeter's probes to different points along the same segments of wire?

**B** Explain why it would be incorrect to define resistance as the amount of charge the resistor allows to flow.

## 3.5 Current-conducting properties of materials

Ohm's law has a remarkable property, which is that current will flow even in response to a voltage difference that is as small as we care to make it. In the analogy of pushing a crate across a floor, it is as though even a flea could slide the crate across the floor, albeit at some very low speed. The flea cannot do this because of static friction, which we can think of as an effect arising from the tendency of the microscopic bumps and valleys in the crate and floor to lock together. The fact that Ohm's law holds for nearly all solids has an interesting interpretation: at least some of the electrons are not "locked down" at all to any specific atom.

More generally we can ask how charge actually flows in various solids, liquids, and gases. This will lead us to the explanations of many interesting phenomena, including lightning, the bluish crust that builds up on the terminals of car batteries, and the need for electrolytes in sports drinks.

### Solids

In atomic terms, the defining characteristic of a solid is that its atoms are packed together, and the nuclei cannot move very far from their equilibrium positions. It makes sense, then, that electrons, not ions, would be the charge carriers when currents flow in solids. This fact was established experimentally by Tolman and Stewart, in an experiment in which they spun a large coil of wire and then abruptly stopped it. They observed a current in the wire immediately after the coil was stopped, which indicated that charged particles that were not permanently locked to a specific atom had continued to move because of their own inertia, even after the material of the wire in general stopped. The direction of the current showed that it was negatively charged particles that kept moving. The current only lasted for an instant, however; as the negatively charged particles collected at the downstream end of the wire, farther particles were prevented joining them due to their electrical repulsion, as well as the attraction from the upstream end, which was left with a net positive charge. Tolman and Stewart were even able to determine the mass-to-charge ratio of the particles. We need not go into the details of the analysis here, but particles with high mass would be difficult to decelerate, leading to a stronger and longer pulse of current, while particles with high charge would feel stronger electrical forces decelerating them, which would cause a weaker and shorter pulse. The mass-to-charge ratio thus determined was consistent with the $m/q$ of the electron to within the accuracy of the experiment, which essentially established that the particles were electrons.

The fact that only electrons carry current in solids, not ions, has many important implications. For one thing, it explains why wires don't fray or turn to dust after carrying current for a long time. Electrons are very small (perhaps even pointlike), and it is easy to imagine them passing between the cracks among the atoms without creating holes or fractures in the atomic framework. For those who know a little chemistry, it also explains why all the best conductors are on the left side of the periodic table. The elements in that area are the ones that have only a very loose hold on their outermost electrons.

### Gases

The molecules in a gas spend most of their time separated from each other by significant distances, so it is not possible for them to conduct electricity the way solids do, by handing off electrons from atom to atom. It is therefore not surprising that gases are good insulators.

Gases are also usually nonohmic. As opposite charges build up on a stormcloud and the ground below, the voltage difference becomes greater and greater. Zero current flows, however, until finally the voltage reaches a certain threshold and we have an impressive example of what is known as a spark or electrical discharge. If air was ohmic, the current between the cloud and the ground would simply increase steadily as the voltage difference increased, rather than being zero until a threshold was reached. This behavior can be explained as follows. At some point, the electrical forces on the air electrons and nuclei of the air molecules become so strong that electrons are ripped right off of some of the molecules. The electrons then accelerate toward either the cloud or the ground, whichever is positively charged, and the positive ions accelerate the opposite way. As these charge carriers accelerate, they strike and ionize other molecules, which produces a rapidly growing cascade.

### Liquids

Molecules in a liquid are able to slide past each other, so ions as well as electrons can carry currents. Pure water is a poor conductor because the water molecules tend to hold onto their electrons strongly, and there are therefore not many electrons or ions available to move. Water can become quite a good conductor, however, with the addition of even a small amount of certain substances called electrolytes, which are typically salts. For example, if we add table salt, NaCl, to water, the NaCl molecules dissolve into $Na^+$ and $Cl^-$ ions, which can then move and create currents. This is why electric currents can flow among the cells in our bodies: cellular fluid is quite salty. When we sweat, we lose not just water but electrolytes, so dehydration plays havoc with our cells' electrical systems. It is for this reason that electrolytes are included in sports drinks and formulas for rehydrating infants who have diarrhea.

Since current flow in liquids involves entire ions, it is not surprising that we can see physical evidence when it has occurred. For example, after a car battery has been in use for a while, the $H_2SO_4$ battery acid becomes depleted of hydrogen ions, which are the main charge carriers that complete the circuit on the inside of the battery. The leftover $SO_4$ then forms a visible blue crust on the battery posts.

### Speed of currents and electrical signals

When I talk on the phone to my mother in law two thousand miles away, I do not notice any delay while the signal makes its way back and forth. Electrical signals therefore must travel very quickly, but how fast exactly? The answer is rather subtle. For the sake of concreteness, let's restrict ourselves to currents in metals, which consist of electrons.

The electrons themselves are only moving at speeds of perhaps a few thousand miles per hour, and their motion is mostly random thermal motion. This shows that the electrons in my phone cannot possibly be zipping back and forth between California and New York fast enough to carry the signals. Even if their thousand-mile-an-hour motion was organized rather than random, it would still take them many minutes to get there. Realistically, it will take the average electron even longer than that to make the trip. The current in the wire consists only of a slow overall drift, at a speed on the order of a few centimeters per second, superimposed on the more rapid random motion. We can compare this with the slow westward drift in the population of the U.S. If we could make a movie of the motion of all the people in the U.S. from outer space, and could watch it at high speed so that the people appeared to be scurrying around like ants, we would think that the motion was fairly random, and we would not immediately notice the westward drift. Only after many years would we realize that the number of people heading west over the Sierras had exceeded the number going east, so that California increased its share of the country's population.

So why are electrical signals so fast if the average drift speed of electrons is so slow? The answer is that a disturbance in an electrical system can move much more quickly than the charges themselves. It is as though we filled a pipe with golf balls and then inserted an extra ball at one end, causing a ball to fall out at the other end. The force propagated to the other end in a fraction of a second, but the balls themselves only traveled a few centimeters in that time.

Because the reality of current conduction is so complex, we often describe things using mental shortcuts that are technically incorrect. This is OK as long as we know that they are just shortcuts. For example, suppose the presidents of France and Russia shake hands, and the French politician has inadvertently picked up a positive electrical charge, which shocks the Russian. We may say that the excess positively charged particles in the French leader's body, which all repel each other, take the handshake as an opportunity to get farther apart by spreading out into two bodies rather than one. In reality, it would be a matter of minutes before the ions in one person's body could actually drift deep into the other's. What really happens is that throughout the body of the recipient of the shock there are already various positive and negative ions which are free to move. Even before the perpetrator's charged hand touches the victim's sweaty palm, the charges in the shocker's body begin to repel the positive ions and attract the negative ions in the other person. The split-second sensation of shock is caused by the sudden jumping of the victim's ions by distances of perhaps a micrometer, this effect occurring simultaneously throughout the whole body, although more violently in the hand and arm, which are closer to the other person.

## 3.6 $\int$ Applications of Calculus

As discussed in example 1 on page 80, the definition of current as the rate of change of charge with respect to time must be reexpressed as a derivative in the case where the rate of change is not constant,

$$I = \frac{dq}{dt} \qquad .$$

*Finding current given charge* *example 8*

▷ A charged balloon falls to the ground, and its charge begins leaking off to the Earth. Suppose that the charge on the balloon is given by $q = ae^{-bt}$. Find the current as a function of time, and interpret the answer.

▷ Taking the derivative, we have

$$I = \frac{dq}{dt} \\ = -abe^{-bt}$$

An exponential function approaches zero as the exponent gets more and more negative. This means that both the charge and the current are decreasing in magnitude with time. It makes sense that the charge approaches zero, since the balloon is losing its charge. It also makes sense that the current is decreasing in magnitude, since charge cannot flow at the same rate forever without overshooting zero.

*Finding charge given current* *example 9*

▷ In the segment of the ATLAS accelerator shown in figure i on page 90, the current flowing back and forth between the two cylinders is given by $I = a\cos bt$. What is the charge on one of the cylinders as a function of time? ▷ We are given the current and want to find the charge, i.e. we are given the derivative and we want to find the original function that would give that derivative. This means we need to integrate:

$$
\begin{aligned}
q &= \int I \mathrm{d}t \\
&= \int a\cos bt \,\mathrm{d}t \\
&= \frac{a}{b}\sin bt + q_o \qquad ,
\end{aligned}
$$

where $q_o$ is a constant of integration.

We can interpret this in order to explain why a superconductor needs to be used. The constant $b$ must be very large, since the current is supposed to oscillate back and forth millions of times a second. Looking at the final result, we see that if $b$ is a very large number, and $q$ is to be a significant amount of charge, then $a$ must be a very large number as well. If $a$ is numerically large, then the current must be very large, so it would heat the accelerator too much if it was flowing through an ordinary conductor.

## Summary

### Selected Vocabulary

| | |
|---|---|
| current . . . . . . | the rate at which charge crosses a certain boundary |
| ampere . . . . . . | the metric unit of current, one coulomb pe second; also "amp" |
| ammeter . . . . . | a device for measuring electrical current |
| circuit . . . . . . . | an electrical device in which charge can come back to its starting point and be recycled rather than getting stuck in a dead end |
| open circuit . . . | a circuit that does not function because it has a gap in it |
| short circuit . . . | a circuit that does not function because charge is given a low-resistance "shortcut" path that it can follow, instead of the path that makes it do something useful |
| voltage . . . . . . | electrical potential energy per unit charge that will be possessed by a charged particle at a certain point in space |
| volt . . . . . . . . | the metric unit of voltage, one joule per coulomb |
| voltmeter-SHARED . . . . | a device for measuring voltage differences |
| ohmic . . . . . . . | describes a substance in which the flow of current between two points is proportional to the voltage difference between them |
| resistance . . . . | the ratio of the voltage difference to the current in an object made of an ohmic substance |
| ohm . . . . . . . . | the metric unit of electrical resistance, one volt per ampere |

### Notation

| | |
|---|---|
| $I$ . . . . . . . . . . | current |
| $A$ . . . . . . . . . | units of amperes |
| $V$ . . . . . . . . . | voltage |
| $V$ . . . . . . . . . | units of volts |
| $R$ . . . . . . . . . | resistance |
| $\Omega$ . . . . . . . . . | units of ohms |

### Other Terminology and Notation

| | |
|---|---|
| electric potential | rather than the more informal "voltage" used here; despite the misleading name, it is not the same as electric potential energy |
| eV . . . . . . . . . | a unit of energy, equal to $e$ multiplied by 1 volt; $1.6 \times 10^{-19}$ joules |

### Summary

All electrical phenomena are alike in that that arise from the presence or motion of charge. Most practical electrical devices are based on the motion of charge around a complete circuit, so that the charge can be recycled and does not hit any dead ends. The most useful measure of the flow of charge is current, $I = \Delta q/\Delta t$.

An electrical device whose job is to transform energy from one form into another, e.g., a lightbulb, uses power at a rate which depends both on how rapidly charge is flowing through it and on how much work is done on each unit of charge. The latter quantity is known as the voltage difference between the point where the current enters the device and the point where the current leaves it. Since there is a type of potential energy associated with electrical forces, the amount of work they do is equal to the difference in potential energy between the two points, and we therefore define voltage differences directly in terms of potential energy, $\Delta V = \Delta PE_{elec}/q$. The rate of power dissipation is $P = I\Delta V$.

Many important electrical phenomena can only be explained if we understand the mechanisms of current flow at the atomic level. In metals, currents are carried by electrons, in liquids by ions. Gases are normally poor conductors unless their atoms are subjected to such intense electrical forces that the atoms become ionized.

Many substances, including all solids, respond to electrical forces in such a way that the flow of current between two points is proportional to the voltage difference between those points. Such a substance is called ohmic, and an object made out of an ohmic substance can be rated in terms of its resistance, $R = \Delta$V/I. An important corollary is that a perfect conductor, with $R = 0$, must have constant voltage everywhere within it.

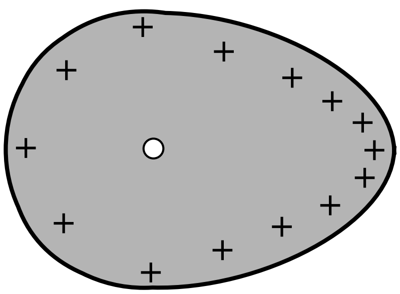
## Problems

### Key

- √ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ⋆ A difficult problem.

**1** A resistor has a voltage difference $\Delta V$ across it, causing a current $I$ to flow.
(a) Find an equation for the power it dissipates as heat in terms of the variables $I$ and $R$ only, eliminating $\Delta V$. √
(b) If an electrical line coming to your house is to carry a given amount of current, interpret your equation from part a to explain whether the wire's resistance should be small, or large. √

**2** (a) Express the power dissipated by a resistor in terms of $R$ and $\Delta V$ only, eliminating $I$. √
(b) Electrical receptacles in your home are mostly 110 V, but circuits for electric stoves, air conditioners, and washers and driers are usually 220 V. The two types of circuits have differently shaped receptacles. Suppose you rewire the plug of a drier so that it can be plugged in to a 110 V receptacle. The resistor that forms the heating element of the drier would normally draw 200 W. How much power does it actually draw now? √

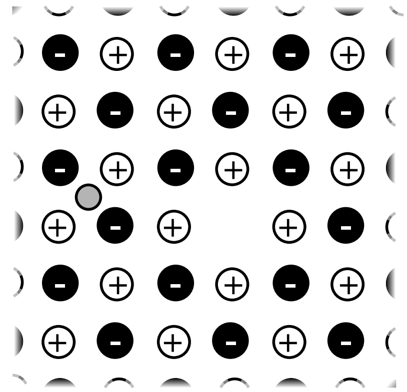


Problem 3.

**3** As discussed in the text, when a conductor reaches an equilibrium where its charge is at rest, there is always zero electric force on a charge in its interior, and any excess charge concentrates itself on the surface. The surface layer of charge arranges itself so as to produce zero total force at any point in the interior. (Otherwise the free charge in the interior could not be at rest.) Suppose you have a teardrop-shaped conductor like the one shown in the figure. Since the teardrop is a conductor, there are free charges everywhere inside it, but consider a free charged particle at the location shown with a white circle. Explain why, in order to produce zero force on this particle, the surface layer of charge must be denser in the pointed part of the teardrop. (Similar reasoning shows why lightning rods are made with points. The charged stormclouds induce positive and negative charges to move to opposite ends of the rod. At the pointed upper end of the rod, the charge tends to concentrate at the point, and this charge attracts the lightning.)

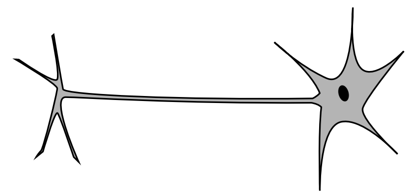
**4** Use the result of problem 3 on page 38 to find an equation for the voltage at a point in space at a distance $r$ from a point charge $Q$. (Take your $V = 0$ distance to be anywhere you like.) √

**5** Referring back to problem 6 on page 39 about the sodium chloride crystal, suppose the lithium ion is going to jump from the gap it is occupying to one of the four closest neighboring gaps. Which one will it jump to, and if it starts from rest, how fast will it be going by the time it gets there? (It will keep on moving and accelerating after that, but that does not concern us.) [Hint: The approach is similar to the one used for the other problem, but you want to work with voltage and potential energy rather than force.] √ ⋆



Problem 5.

**6** Referring back to our old friend the neuron from problem 1 on page 38, let's now consider what happens when the nerve is stimulated to transmit information. When the blob at the top (the cell body) is stimulated, it causes $Na^+$ ions to rush into the top of the tail (axon). This electrical pulse will then travel down the axon, like a flame burning down from the end of a fuse, with the $Na^+$ ions at each point first going out and then coming back in. If $10^{10}$ $Na^+$ ions cross the cell membrane in 0.5 ms, what amount of current is created? √



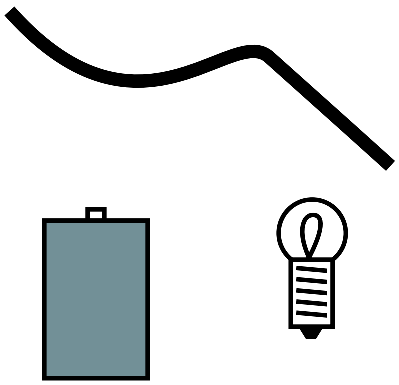
Problem 6.

**7** If a typical light bulb draws about 900 mA from a 110-V household circuit, what is its resistance? (Don't worry about the fact that it's alternating current.) √

**8** Today, even a big luxury car like a Cadillac can have an electrical system that is relatively low in power, since it doesn't need to do much more than run headlights, power windows, etc. In the near future, however, manufacturers plan to start making cars with electrical systems about five times more powerful. This will allow certain energy-wasting parts like the water pump to be run on electrical motors and turned off when they're not needed — currently they're run directly on shafts from the motor, so they can't be shut off. It may even be possible to make an engine that can shut off at a stoplight and then turn back on again without cranking, since the valves can be electrically powered. Current cars' electrical systems have 12-volt batteries (with 14-volt chargers), but the new systems will have 36-volt batteries (with 42-volt chargers).
(a) Suppose the battery in a new car is used to run a device that requires the same amount of power as the corresponding device in the old car. Based on the sample figures above, how would the currents handled by the wires in one of the new cars compare with the currents in the old ones?
(b) The real purpose of the greater voltage is to handle devices that need *more* power. Can you guess why they decided to change to 36-volt batteries rather than increasing the power without increasing the voltage? √

**9** (a) You take an LP record out of its sleeve, and it acquires a static charge of 1 nC. You play it at the normal speed of $33\frac{1}{3}$ r.p.m., and the charge moving in a circle creates an electric current. What is the current, in amperes? √

(b) Although the planetary model of the atom can be made to work with any value for the radius of the electrons' orbits, more advanced models that we will study later in this course predict definite radii. If the electron is imagined as circling around the proton at a speed of $2.2\times10^6$ m/s, in an orbit with a radius of 0.05 nm, what electric current is created? √ ⋆

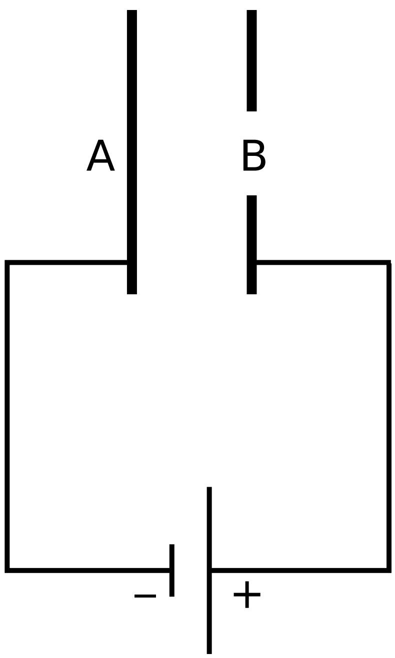


Problem 11.

**10** We have referred to resistors *dissipating* heat, i.e. we have assumed that $P = I\Delta V$ is always greater than zero. Could $I\Delta V$ come out to be negative for a resistor? If so, could one make a refrigerator by hooking up a resistor in such a way that it absorbed heat instead of dissipating it?

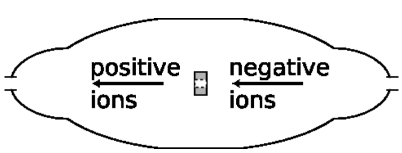
**11** You are given a battery, a flashlight bulb, and a single piece of wire. Draw at least two configurations of these items that would result in lighting up the bulb, and at least two that would not light it. (Don't draw schematics.) If you're not sure what's going on, borrow the materials from your instructor and try it. Note that the bulb has two electrical contacts: one is the threaded metal jacket, and the other is the tip. [Problem by Arnold Arons.]

**12** In a wire carrying a current of 1.0 pA, how long do you have to wait, on the average, for the next electron to pass a given point? Express your answer in units of microseconds.
▷ Solution, p. 196




Problem 13.

**13** The figure shows a simplified diagram of an electron gun such as the one used in the Thomson experiment, or the one that creates the electron beam in a TV tube. Electrons that spontaneously emerge from the negative electrode (cathode) are then accelerated to the positive electrode, which has a hole in it. (Once they emerge through the hole, they will slow down. However, if the two electrodes are fairly close together, this slowing down is a small effect, because the attractive and repulsive forces experienced by the electron tend to cancel.) (a) If the voltage difference between the electrodes is ΔV, what is the velocity of an electron as it emerges at B? (Assume its initial velocity, at A, is negiligible.) (b) Evaluate your expression numerically for the case where $\Delta V$=10 kV, and compare to the speed of light.
▷ Solution, p. 197

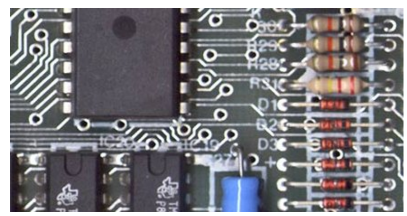



Problem 14.

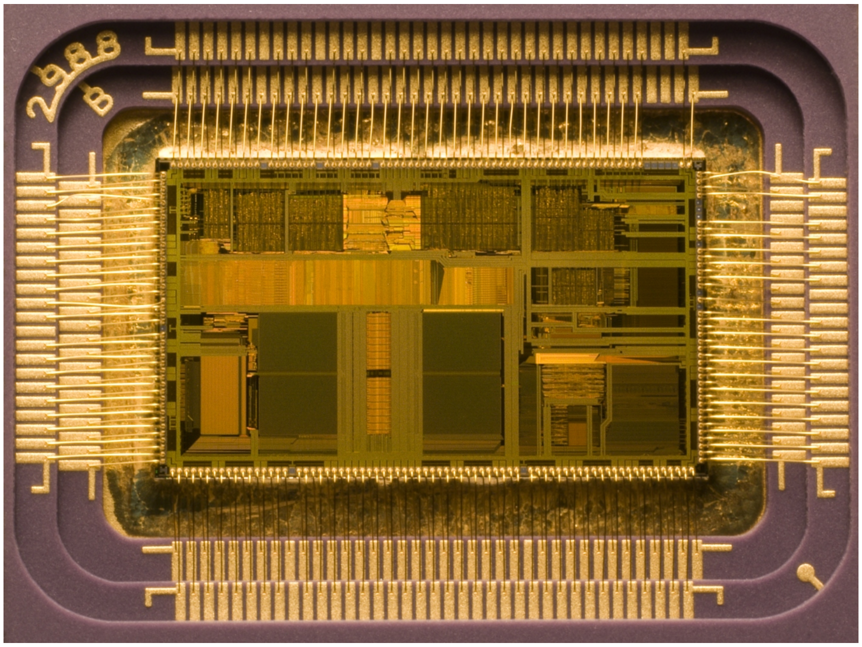
**14** The figure shows a simplified diagram of a device called a tandem accelerator, used for accelerating beams of ions up to speeds on the order of 1% of the speed of light. The nuclei of these ions collide with the nuclei of atoms in a target, producing nuclear reactions for experiments studying the structure of nuclei. The outer shell of the accelerator is a conductor at zero voltage (i.e., the same voltage as the Earth). The electrode at the center, known as the "terminal," is at a high positive voltage, perhaps millions of volts. Negative ions with a charge of −1 unit (i.e., atoms with one extra electron) are produced offstage on the right, typically by chemical reactions with cesium, which is a chemical element that has a strong tendency to give away electrons. Relatively weak electric and magnetic forces are used to transport these $-1$ ions into the accelerator, where they are attracted to the terminal. Although the center of the terminal has a hole in it to let the ions pass through, there is a very thin carbon foil there that they must physically penetrate. Passing through the foil strips off some number of electrons, changing the atom into a positive ion, with a charge of $+n$ times the fundamental charge. Now that the atom is positive, it is repelled by the terminal, and accelerates some more on its way out of the accelerator. (a) Find the velocity, $v$, of the emerging beam of positive ions, in terms of $n$, their mass $m$, the terminal voltage $V$, and fundamental constants. Neglect the small change in mass caused by the loss of electrons in the stripper foil. (b) To fuse protons with protons, a minimum beam velocity of about 11% of the speed of light is required. What terminal voltage would be needed in this case?

**15** Three charges, each of strength $Q (Q > 0)$ form a fixed equilateral triangle with sides of length $b$. You throw a particle of mass $m$ and positive charge $q$ from far away, with an initial speed $v$. Your goal is to get the particle to go to the center of the triangle, your aim is perfect, and you are free to throw from any direction you like. What is the minimum possible value of $v$?

**16** You have to do different things with a circuit to measure current than to measure a voltage difference. Which would be more practical for a printed circuit board, in which the wires are actually strips of metal embedded inside the board? ▷ Solution, p. 197



A printed circuit board, like the kind referred to in problem 16.



An Intel 486 computer chip in its packaging.

# Chapter 4

# Circuits, Part 2

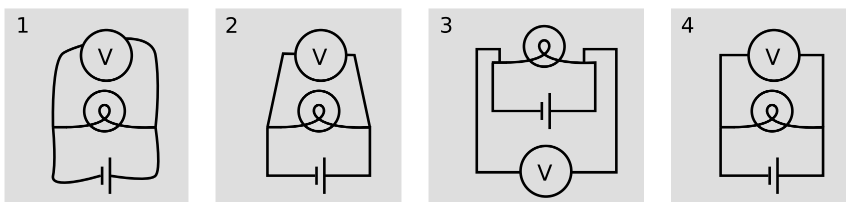
In chapter 3 we limited ourselves to relatively simple circuits, essentially nothing more than a battery and a single lightbulb. The purpose of this chapter is to introduce you to more complex circuits, containing multiple resistors or voltage sources in series, in parallel, or both.

Why do you need to know this stuff? After all, if you were planning on being an electrical engineer you probably wouldn't be learning physics from this book. Consider, however, that every time you plug in a lamp or a radio you are adding a circuit element to a household circuit and making it more complex. Electrical safety, as well, cannot really be understood without understanding multiple-element circuits, since getting shocked usually involves at least two parts: the device that is supposed to use power plus the body of the person in danger. If you are a student majoring in the life sciences, you should realize as well that all cells are inherently electrical, and in any multicellular organism there will therefore be various series and parallel circuits.

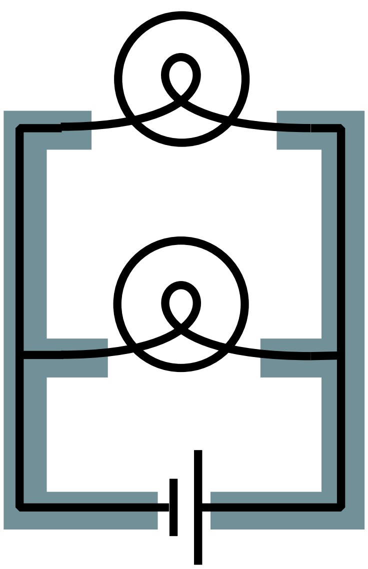
Even apart from these practical purposes, there is a very fundamental reason for reading this chapter: to understand chapter 3 better. At this point in their studies, I always observe students using words and modes of reasoning that show they have not yet become completely comfortable and fluent with the concepts of voltage and current. They ask, "aren't voltage and current sort of the same idea?" They speak of voltage "going through" a lightbulb. Once they begin honing their skills on more complicated circuits I always see their confidence and understanding increase immeasurably.

## 4.1 Schematics

I see a chess position; Kasparov sees an interesting Ruy Lopez variation. To the uninitiated a schematic may look as unintelligible as Mayan hieroglyphs, but even a little bit of eye training can go a long way toward making its meaning leap off the page. A schematic is a stylized and simplified drawing of a circuit. The purpose is to eliminate as many irrelevant features as possible, so that the relevant ones are easier to pick out.




a / 1. Wrong: The shapes of the wires are irrelevant. 2. Wrong: Right angles should be used. 3. Wrong: A simple pattern is made to look unfamiliar and complicated. 4. Right.



b / The two shaded areas shaped like the letter "E" are both regions of constant voltage.

An example of an irrelevant feature is the physical shape, length, and diameter of a wire. In nearly all circuits, it is a good approximation to assume that the wires are perfect conductors, so that any piece of wire uninterrupted by other components has constant voltage throughout it. Changing the length of the wire, for instance, does not change this fact. (Of course if we used miles and miles of wire, as in a telephone line, the wire's resistance would start to add up, and its length would start to matter.) The shapes of the wires are likewise irrelevant, so we draw them with standardized, stylized shapes made only of vertical and horizontal lines with right-angle bends in them. This has the effect of making similar circuits look more alike and helping us to recognize familiar patterns, just as words in a newspaper are easier to recognize than handwritten ones. Figure a shows some examples of these concepts.

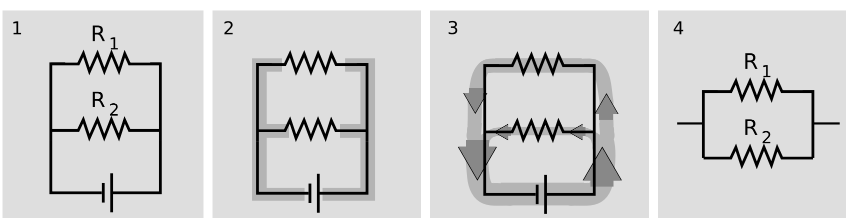
The most important first step in learning to read schematics is to learn to recognize contiguous pieces of wire which must have constant voltage throughout. In figure b, for example, the two shaded E-shaped pieces of wire must each have constant voltage. This focuses our attention on two of the main unknowns we'd like to be able to predict: the voltage of the left-hand E and the voltage of the one on the right.

## 4.2 Parallel resistances and the junction rule

One of the simplest examples to analyze is the parallel resistance circuit, of which figure b was an example. In general we may have unequal resistances $R_1$ and $R_2$, as in c/1. Since there are only two constant-voltage areas in the circuit, c/2, all three components have the same voltage difference across them. A battery normally succeeds in maintaining the voltage differences across itself for which it was designed, so the voltage drops $\Delta V_1$ and $\Delta V_2$ across the resistors must both equal the voltage of the battery:

$$\Delta V_1 = \Delta V_2 = \Delta V_{battery} \qquad .$$

Each resistance thus feels the same voltage difference as if it was the only one in the circuit, and Ohm's law tells us that the amount of current flowing through each one is also the same as it would have been in a one-resistor circuit. This is why household electrical circuits are wired in parallel. We want every appliance to work the same, regardless of whether other appliances are plugged in or unplugged, turned on or switched off. (The electric company doesn't use batteries of course, but our analysis would be the same for any device that maintains a constant voltage.)




c / 1. Two resistors in parallel. 2. There are two constant-voltage areas. 3. The current that comes out of the battery splits between the two resistors, and later reunites. 4. The two resistors in parallel can be treated as a single resistor with a smaller resistance value.

Of course the electric company can tell when we turn on every light in the house. How do they know? The answer is that we draw more current. Each resistance draws a certain amount of current, and the amount that has to be supplied is the sum of the two individual currents. The current is like a river that splits in half, c/3, and then reunites. The total current is

$$I_{total} = I_1 + I_2 \qquad .$$

This is an example of a general fact called the junction rule:

**the junction rule**
In any circuit that is not storing or releasing charge, conservation of charge implies that the total current flowing out of any junction must be the same as the total flowing in.

Coming back to the analysis of our circuit, we apply Ohm's law to each resistance, resulting in

$$
\begin{aligned}
I_{total} &= \Delta V/R_1 + \Delta V/R_2 \\
&= \Delta V \left( \frac{1}{R_1} + \frac{1}{R_2} \right) \qquad .
\end{aligned}
$$

As far as the electric company is concerned, your whole house is just one resistor with some resistance $R$, called the *equivalent resistance*. They would write Ohm's law as

$$
I_{total} = \Delta V/R \qquad ,
$$

from which we can determine the equivalent resistance by comparison with the previous expression:

$$
\begin{aligned}
1/R &= \frac{1}{R_1} + \frac{1}{R_2} \\
R &= \left( \frac{1}{R_1} + \frac{1}{R_2} \right)^{-1}
\end{aligned}
$$

[equivalent resistance of two resistors in parallel]

Two resistors in parallel, c/4, are equivalent to a single resistor with a value given by the above equation.

*Two lamps on the same household circuit* *example 1*

▷ You turn on two lamps that are on the same household circuit. Each one has a resistance of 1 ohm. What is the equivalent resistance, and how does the power dissipation compare with the case of a single lamp?

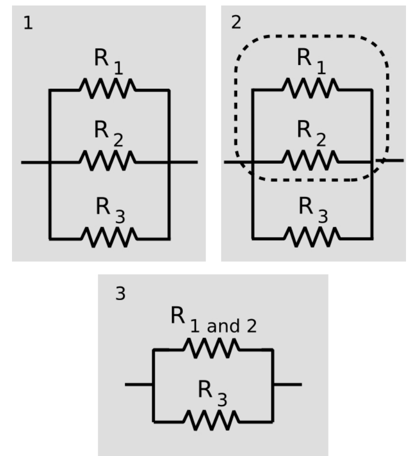
▷ The equivalent resistance of the two lamps in parallel is

$$
\begin{aligned}
R &= \left( \frac{1}{R_1} + \frac{1}{R_2} \right)^{-1} \\
&= \left( \frac{1}{1\ \Omega} + \frac{1}{1\ \Omega} \right)^{-1} \\
&= \left( 1\ \Omega^{-1} + 1\ \Omega^{-1} \right)^{-1} \\
&= \left( 2\ \Omega^{-1} \right)^{-1} \\
&= 0.5\ \Omega
\end{aligned}
$$

The voltage difference across the whole circuit is always the 110 V set by the electric company (it's alternating current, but that's irrelevant). The resistance of the whole circuit has been cut in half by turning on the second lamp, so a fixed amount of voltage will produce twice as much current. Twice the current flowing across the same voltage difference means twice as much power dissipation, which makes sense.

The cutting in half of the resistance surprises many students, since we are "adding more resistance" to the circuit by putting in the second lamp. Why does the equivalent resistance come out to be less than the resistance of a single lamp? This is a case where purely verbal reasoning can be misleading. A resistive circuit element, such as the filament of a lightbulb, is neither a perfect insulator nor a perfect conductor. Instead of analyzing this type of circuit in terms of "resistors," i.e., partial insulators, we could have spoken of "conductors." This example would then seem reasonable, since we "added more conductance," but one would then have the incorrect expectation about the case of resistors in series, discussed in the following section.

Perhaps a more productive way of thinking about it is to use mechanical intuition. By analogy, your nostrils resist the flow of air through them, but having two nostrils makes it twice as easy to breathe.



d / Three resistors in parallel.

*Three resistors in parallel* *example 2*

▷ What happens if we have three or more resistors in parallel?

▷ This is an important example, because the solution involves an important technique for understanding circuits: breaking them down into smaller parts and them simplifying those parts. In the circuit d/1, with three resistors in parallel, we can think of two of the resistors as forming a single resistor, d/2, with equivalent resistance

$$R_{12} = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{-1} \quad .$$

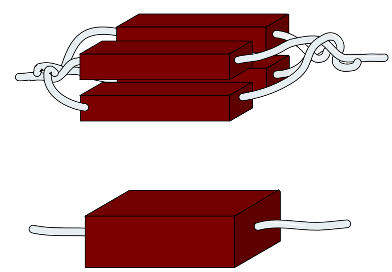
We can then simplify the circuit as shown in d/3, so that it contains only two resistances. The equivalent resistance of the whole circuit is then given by

$$R_{123} = \left(\frac{1}{R_{12}} + \frac{1}{R_3}\right)^{-1} \quad .$$

Substituting for $R_{12}$ and simplifying, we find the result

$$R_{123} = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right)^{-1} \quad ,$$

which you probably could have guessed. The interesting point here is the divide-and-conquer concept, not the mathematical result.


e / Uniting four resistors in parallel is equivalent to making a single resistor with the same length but four times the cross-sectional area. The result is to make a resistor with one quarter the resistance.

*An arbitrary number of identical resistors in parallel* *example 3*

▷ What is the resistance of $N$ identical resistors in parallel?

▷ Generalizing the results for two and three resistors, we have

$$R_N = \left(\frac{1}{R_1} + \frac{1}{R_2} + \ldots\right)^{-1} \quad ,$$

where "..." means that the sum includes all the resistors. If all the resistors are identical, this becomes

$$R_N = \left(\frac{N}{R}\right)^{-1} = \frac{R}{N}$$

*Dependence of resistance on cross-sectional area* *example 4*

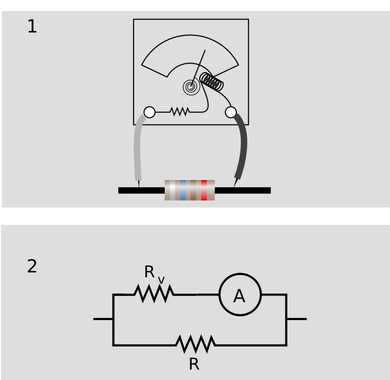
We have alluded briefly to the fact that an object's electrical resistance depends on its size and shape, but now we are ready to begin making more mathematical statements about it. As suggested by figure e, increasing a resistors's cross-sectional area is equivalent to adding more resistors in parallel, which will lead to an overall decrease in resistance. Any real resistor with straight, parallel sides can be sliced up into a large number of pieces, each with cross-sectional area of, say, 1 $\mu m^2$. The number, $N$, of such slices is proportional to the total cross-sectional area of the resistor, and by application of the result of the previous example we therefore find that the resistance of an object is inversely proportional to its cross-sectional area.

f / A fat pipe has less resistance than a skinny pipe.

An analogous relationship holds for water pipes, which is why high-flow trunk lines have to have large cross-sectional areas. To make lots of water (current) flow through a skinny pipe, we'd need an impractically large pressure (voltage) difference.

*Incorrect readings from a voltmeter* *example 5*

A voltmeter is really just an ammeter with an internal resistor, and we use a voltmeter in parallel with the thing that we're trying to measure the voltage difference across. This means that any time we measure the voltage drop across a resistor, we're essentially putting two resistors in parallel. The ammeter inside the voltmeter can be ignored for the purpose of analyzing what how current flows in the circuit, since it is essentially just some coiled-up wire with a very low resistance.




g / A voltmeter is really an ammeter with an internal resistor. When we measure the voltage difference across a resistor, 1, we are really constructing a parallel resistance circuit, 2.

Now if we are carrying out this measurement on a resistor that is part of a larger circuit, we have changed the behavior of the circuit through our act of measuring. It is as though we had modified the circuit by replacing the resistance $R$ with the smaller equivalent resistance of $R$ and $R_v$ in parallel. It is for this reason that voltmeters are built with the largest possible internal resistance. As a numerical example, if we use a voltmeter with an internal resistance of 1 $M\Omega$ to measure the voltage drop across a one-ohm resistor, the equivalent resistance is 0.999999 $\Omega$, which is not different enough to make any difference. But if we tried to use the same voltmeter to measure the voltage drop across a $2-M\Omega$ resistor, we would be reducing the resistance of that part of the circuit by a factor of three, which would produce a drastic change in the behavior of the whole circuit.

This is the reason why you can't use a voltmeter to measure the voltage difference between two different points in mid-air, or between the ends of a piece of wood. This is by no means a stupid thing to want to do, since the world around us is not a constant-voltage environment, the most extreme example being when an electrical storm is brewing. But it will not work with an ordinary voltmeter because the resistance of the air or the wood is many gigaohms. The effect of waving a pair of voltmeter probes around in the air is that we provide a reuniting path for the positive and negative charges that have been separated — through the voltmeter itself, which is a good conductor compared to the air. This reduces to zero the voltage difference we were trying to measure.

In general, a voltmeter that has been set up with an open circuit (or a very large resistance) between its probes is said to be "floating." An old-fashioned analog voltmeter of the type described here will read zero when left floating, the same as when it was sitting on the shelf. A floating digital voltmeter usually shows an error message.

## 4.3 Series resistances

The two basic circuit layouts are parallel and series, so a pair of resistors in series, h/1, is another of the most basic circuits we can make. By conservation of charge, all the current that flows through one resistor must also flow through the other (as well as through the battery):

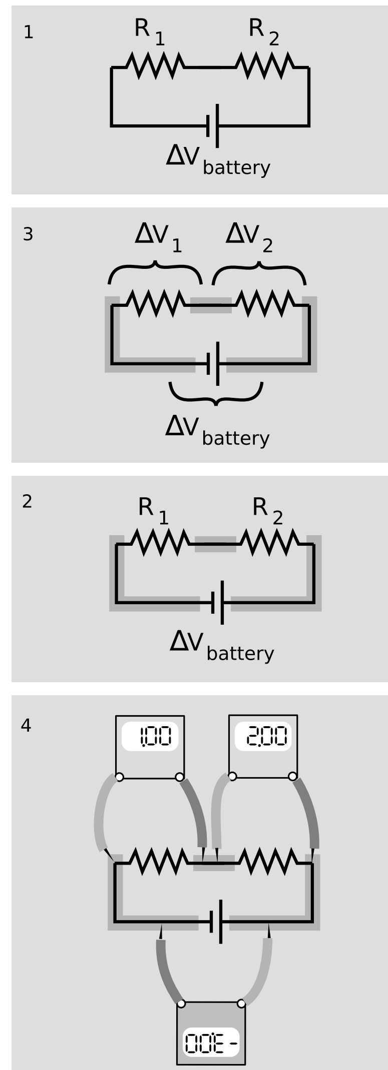
$$I_1 = I_2 \qquad .$$

The only way the information about the two resistance values is going to be useful is if we can apply Ohm's law, which will relate the resistance of each resistor to the current flowing through it and the voltage difference across it. Figure h/2 shows the three constant-voltage areas. Voltage differences are more physically significant than voltages, so we define symbols for the voltage differences across the two resistors in figure h/3.

We have three constant-voltage areas, with symbols for the difference in voltage between every possible pair of them. These three voltage differences must be related to each other. It is as though I tell you that Fred is a foot taller than Ginger, Ginger is a foot taller than Sally, and Fred is two feet taller than Sally. The information is redundant, and you really only needed two of the three pieces of data to infer the third. In the case of our voltage differences, we have

$$|\Delta V_1| + |\Delta V_2| = |\Delta V_{battery}| \qquad .$$

The absolute value signs are because of the ambiguity in how we define our voltage differences. If we reversed the two probes of the voltmeter, we would get a result with the opposite sign. Digital voltmeters will actually provide a minus sign on the screen if the




h / 1. A battery drives current through two resistors in series. 2. There are three constant-voltage regions. 3. The three voltage differences are related. 4. If the meter crab-walks around the circuit without flipping over or crossing its legs, the resulting voltages have plus and minus signs that make them add up to zero.

wire connected to the "V" plug is lower in voltage than the one connected to the "COM" plug. Analog voltmeters pin the needle against a peg if you try to use them to measure negative voltages, so you have to fiddle to get the leads connected the right way, and then supply any necessary minus sign yourself.

Figure h/4 shows a standard way of taking care of the ambiguity in signs. For each of the three voltage measurements around the loop, we keep the same probe (the darker one) on the clockwise side. It is as though the voltmeter was sidling around the circuit like a crab, without ever "crossing its legs." With this convention, the relationship among the voltage drops becomes

$$\Delta V_1 + \Delta V_2 = -\Delta V_{battery} \qquad ,$$

or, in more symmetrical form,

$$\Delta V_1 + \Delta V_2 + \Delta V_{battery} = 0 \qquad .$$

More generally, this is known as the loop rule for analyzing circuits:

**the loop rule**
Assuming the standard convention for plus and minus signs, the sum of the voltage drops around any closed loop in a circuit must be zero.

Looking for an exception to the loop rule would be like asking for a hike that would be downhill all the way and that would come back to its starting point!

For the circuit we set out to analyze, the equation

$$\Delta V_1 + \Delta V_2 + \Delta V_{battery} = 0$$

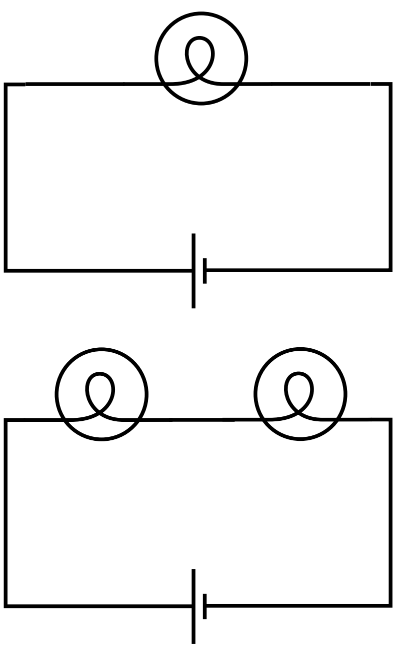
can now be rewritten by applying Ohm's law to each resistor:

$$I_1 R_1 + I_2 R_2 + \Delta V_{battery} = 0 \qquad .$$

The currents are the same, so we can factor them out:

$$I\left(R_1 + R_2\right) + \Delta V_{battery} = 0 \qquad ,$$

and this is the same result we would have gotten if we had been analyzing a one-resistor circuit with resistance $R_1 + R_2$. Thus the equivalent resistance of resistors in series equals the sum of their resistances.



i / Example 6.

*Two lightbulbs in series* *example 6*

▷ If two identical lightbulbs are placed in series, how do their brightnesses compare with the brightness of a single bulb?

▷ Taken as a whole, the pair of bulbs act like a doubled resistance, so they will draw half as much current from the wall. Each bulb will be dimmer than a single bulb would have been.

The total power dissipated by the circuit is $I\Delta V$. The voltage drop across the whole circuit is the same as before, but the current is halved, so the two-bulb circuit draws half as much total power as the one-bulb circuit. Each bulb draws one-quarter of the normal power.

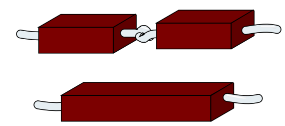
Roughly speaking, we might expect this to result in one quarter the light being produced by each bulb, but in reality lightbulbs waste quite a high percentage of their power in the form of heat and wavelengths of light that are not visible (infrared and ultraviolet). Less light will be produced, but it's hard to predict exactly how much less, since the efficiency of the bulbs will be changed by operating them under different conditions.

*More than two equal resistances in series* — *example 7*

By straightforward application of the divide-and-conquer technique discussed in the previous section, we find that the equivalent resistance of $N$ identical resistances $R$ in series will be $NR$.

*Dependence of resistance on length* — *example 8*

In the previous section, we proved that resistance is inversely proportional to cross-sectional area. By equivalent reason about resistances in series, we find that resistance is proportional to length. Analogously, it is harder to blow through a long straw than through a short one.



j / Doubling the length of a resistor is like putting two resistors in series. The resistance is doubled.

Putting the two arguments together, we find that the resistance of an object with straight, parallel sides is given by

$$R = (\text{constant}) \cdot L/A$$

The proportionality constant is called the resistivity, and it depends only on the substance of which the object is made. A resistivity measurement could be used, for instance, to help identify a sample of an unknown substance.

*Choice of high voltage for power lines* — *example 9*

Thomas Edison got involved in a famous technological controversy over the voltage difference that should be used for electrical power lines. At this time, the public was unfamiliar with electricity, and easily scared by it. The president of the United States, for instance, refused to have electrical lighting in the White House when it first became commercially available because he considered it unsafe, preferring the known fire hazard of oil lamps to the mysterious dangers of electricity. Mainly as a way to overcome public fear, Edison believed that power should be transmitted using small voltages, and he publicized his opinion by giving demonstrations at which a dog was lured into position to be killed by a large voltage difference between two sheets of metal on the ground. (Edison's opponents also advocated alternating current rather than direct current, and AC is more dangerous than DC as well. As we will discuss later, AC can be easily stepped up and down to the desired voltage level using a device called a transformer.)

Now if we want to deliver a certain amount of power $P_L$ to a load such as an electric lightbulb, we are constrained only by the equation $P_L = I\Delta V_L$. We can deliver any amount of power we wish, even with a low voltage, if we are willing to use large currents. Modern electrical distribution networks, however, use dangerously high voltage differences of tens of thousands of volts. Why did Edison lose the debate?

It boils down to money. The electric company must deliver the amount of power $P_L$ desired by the customer through a transmission line whose resistance $R_T$ is fixed by economics and geography. The same current flows through both the load and the transmission line, dissipating power usefully in the former and wastefully in the latter. The efficiency of the system is

$$\begin{aligned} \text{efficiency} &= \frac{\text{power paid for by the customer}}{\text{power paid for by the utility}} \\ &= \frac{P_L}{P_L + P_T} \\ &= \frac{1}{1 + P_T/P_L} \end{aligned}$$

Putting ourselves in the shoes of the electric company, we wish to get rid of the variable $P_T$, since it is something we control only indirectly by our choice of $\Delta V_T$ and $I$. Substituting $P_T = I\Delta V_T$, we find

$$\text{efficiency} = \frac{1}{1 + \frac{I\Delta V_T}{P_L}}$$

We assume the transmission line (but not necessarily the load) is ohmic, so substituting $\Delta V_T = IR_T$ gives

$$efficiency = \frac{1}{1 + \frac{I^2 R_T}{P_L}}$$

This quantity can clearly be maximized by making $I$ as small as possible, since we will then be dividing by the smallest possible quantity on the bottom of the fraction. A low-current circuit can only deliver significant amounts of power if it uses high voltages, which is why electrical transmission systems use dangerous high voltages.

*Getting killed by your ammeter* *example 10*

As with a voltmeter, an ammeter can give erroneous readings if it is used in such a way that it changes the behavior the circuit. An ammeter is used in series, so if it is used to measure the current through a resistor, the resistor's value will effectively be changed to $R + R_a$, where $R_a$ is the resistance of the ammeter. Ammeters are designed with very low resistances in order to make it unlikely that $R + R_a$ will be significantly different from $R$.

In fact, the real hazard is death, not a wrong reading! Virtually the only circuits whose resistances are significantly less than that of an ammeter are those designed to carry huge currents. An ammeter inserted in such a circuit can easily melt. When I was working at a laboratory funded by the Department of Energy, we got periodic bulletins from the DOE safety office about serious accidents at other sites, and they held a certain ghoulish fascination. One of these was about a DOE worker who was completely incinerated by the explosion created when he inserted an ordinary Radio Shack ammeter into a high-current circuit. Later estimates showed that the heat was probably so intense that the explosion was a ball of plasma — a gas so hot that its atoms have been ionized.

## Discussion Questions

**A** We have stated the loop rule in a symmetric form where a series of voltage drops adds up to zero. To do this, we had to define a standard way of connecting the voltmeter to the circuit so that the plus and minus signs would come out right. Suppose we wish to restate the junction rule in a similar symmetric way, so that instead of equating the current coming in to the current going out, it simply states that a certain sum of currents at a junction adds up to zero. What standard way of inserting the ammeter would we have to use to make this work?

## Summary

A schematic is a drawing of a circuit that standardizes and stylizes its features to make it easier to understand. Any circuit can be broken down into smaller parts. For instance, one big circuit may be understood as two small circuits in series, another as three circuits in parallel. When circuit elements are combined in parallel and in series, we have two basic rules to guide us in understanding how the parts function as a whole:

**the junction rule:** In any circuit that is not storing or releasing charge, conservation of charge implies that the total current flowing out of any junction must be the same as the total flowing in.

**the loop rule:** Assuming the standard convention for plus and minus signs, the sum of the voltage drops around any closed loop in a circuit must be zero.

The simplest application of these rules is to pairs of resistors combined in series or parallel. In such cases, the pair of resistors acts just like a single unit with a certain resistance value, called their equivalent resistance. Resistances in series add to produce a larger equivalent resistance,

$$R_{series} = R_1 + R_2 \qquad ,$$

because the current has to fight its way through both resistances. Parallel resistors combine to produce an equivalent resistance that is smaller than either individual resistance,

$$R_{parallel} = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{-1} \qquad ,$$

because the current has two different paths open to it.

An important example of resistances in parallel and series is the use of voltmeters and ammeters in resistive circuits. A voltmeter acts as a large resistance in parallel with the resistor across which the voltage drop is being measured. The fact that its resistance is not infinite means that it alters the circuit it is being used to investigate, producing a lower equivalent resistance. An ammeter acts as a small resistance in series with the circuit through which the current is to be determined. Its resistance is not quite zero, which leads to an increase in the resistance of the circuit being tested.

## Problems

### Key

√ A computerized answer check is available online.
∫ A problem that requires calculus.
⋆ A difficult problem.

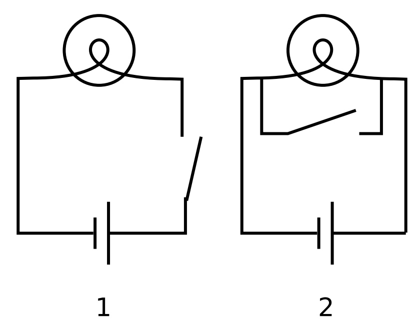
**1** (a) Many battery-operated devices take more than one battery. If you look closely in the battery compartment, you will see that the batteries are wired in series. Consider a flashlight circuit. What does the loop rule tell you about the effect of putting several batteries in series in this way?
(b) The cells of an electric eel's nervous system are not that different from ours — each cell can develop a voltage difference across it of somewhere on the order of one volt. How, then, do you think an electric eel can create voltages of thousands of volts between different parts of its body?

**2** The heating element of an electric stove is connected in series with a switch that opens and closes many times per second. When you turn the knob up for more power, the fraction of the time that the switch is closed increases. Suppose someone suggests a simpler alternative for controlling the power by putting the heating element in series with a variable resistor controlled by the knob. (With the knob turned all the way clockwise, the variable resistor's resistance is nearly zero, and when it's all the way counterclockwise, its resistance is essentially infinite.) (a) Draw schematics. (b) Why would the simpler design be undesirable?

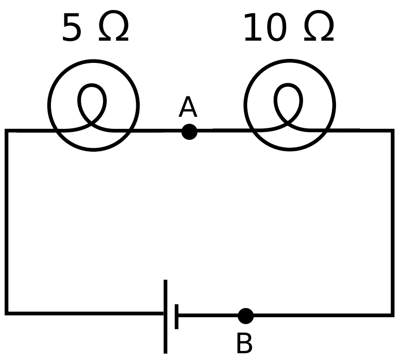
**3** A 1.0 Ω toaster and a 2.0 Ω lamp are connected in parallel with the 110-V supply of your house. (Ignore the fact that the voltage is AC rather than DC.)
(a) Draw a schematic of the circuit.
(b) For each of the three components in the circuit, find the current passing through it and the voltage drop across it.
(c) Suppose they were instead hooked up in series. Draw a schematic and calculate the same things.

**4** Wire is sold in a series of standard diameters, called "gauges." The difference in diameter between one gauge and the next in the series is about 20%. How would the resistance of a given length of wire compare with the resistance of the same length of wire in the next gauge in the series? √

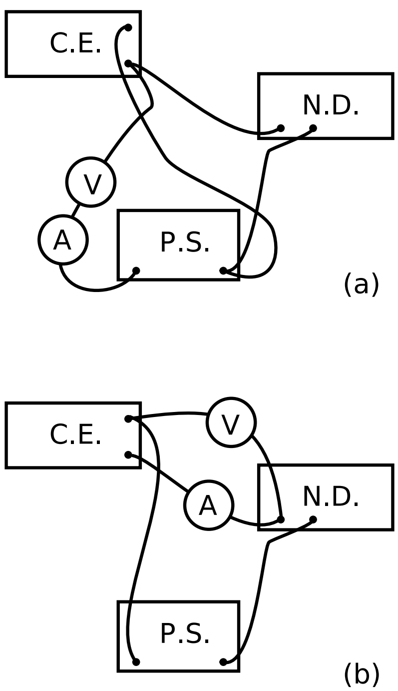
**5** The figure shows two possible ways of wiring a flashlight with a switch. Both will serve to turn the bulb on and off, although the switch functions in the opposite sense. Why is method 1 preferable?




Problem 5.




Problem 6.




Problem 8.

**6** In the figure, the battery is 9 V.
(a) What are the voltage differences across each light bulb?
(b) What current flows through each of the three components of the circuit?
(c) If a new wire is added to connect points A and B, how will the appearances of the bulbs change? What will be the new voltages and currents?
(d) Suppose no wire is connected from A to B, but the two bulbs are switched. How will the results compare with the results from the original setup as drawn?

**7** You have a circuit consisting of two unknown resistors in series, and a second circuit consisting of two unknown resistors in parallel.
(a) What, if anything, would you learn about the resistors in the series circuit by finding that the currents through them were equal?
(b) What if you found out the voltage differences across the resistors in the series circuit were equal?
(c) What would you learn about the resistors in the parallel circuit from knowing that the currents were equal?
(d) What if the voltages in the parallel circuit were equal?

**8** A student in a biology lab is given the following instructions: "Connect the cerebral eraser (C.E.) and the neural depolarizer (N.D.) in parallel with the power supply (P.S.). (Under no circumstances should you ever allow the cerebral eraser to come within 20 cm of your head.) Connect a voltmeter to measure the voltage across the cerebral eraser, and also insert an ammeter in the circuit so that you can make sure you don't put more than 100 mA through the neural depolarizer." The diagrams show two lab groups' attempts to follow the instructions. (a) Translate diagram a into a standard-style schematic. What is correct and incorrect about this group's setup? (b) Do the same for diagram b.

**9** How many different resistance values can be created by combining three unequal resistors? (Don't count possibilities where not all the resistors are used.)

**10** A person in a rural area who has no electricity runs an extremely long extension cord to a friend's house down the road so she can run an electric light. The cord is so long that its resistance, $x$, is not negligible. Show that the lamp's brightness is greatest if its resistance, $y$, is equal to $x$. Explain physically why the lamp is dim for values of $y$ that are too small or too large. $\int$

**11** What resistance values can be created by combining a 1 kΩ resistor and a 10 kΩ resistor? hwsoln:tenkohmonekohm

▷ Solution, p. 197

**12** Suppose six identical resistors, each with resistance $R$, are connected so that they form the edges of a tetrahedron (a pyramid with three sides in addition to the base, i.e., one less side than an Egyptian pyramid). What resistance value or values can be obtained by making connections onto any two points on this arrangement?
▷ Solution, p. 197 ⋆

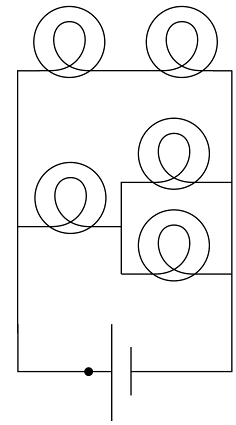
**13** The figure shows a circuit containing five lightbulbs connected to a battery. Suppose you're going to connect one probe of a voltmeter to the circuit at the point marked with a dot. How many unique, nonzero voltage differences could you measure by connecting the other probe to other wires in the circuit?

**14** The lightbulbs in the figure are all identical. If you were inserting an ammeter at various places in the circuit, how many unique currents could you measure? If you know that the current measurement will give the same number in more than one place, only count that as one unique current.

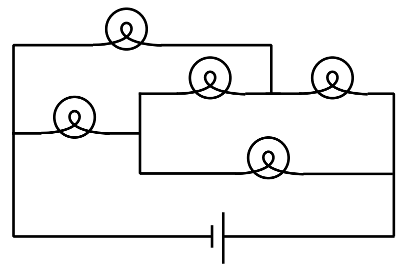
**15** The bulbs are all identical. Which one doesn't light up? ⋆

**16** Each bulb has a resistance of one ohm. How much power is drawn from the one-volt battery? ⋆

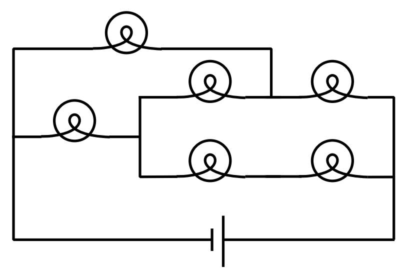
**17** The bulbs all have unequal resistances. Given the three currents shown in the figure, find the currents through bulbs A, B, C, and D.



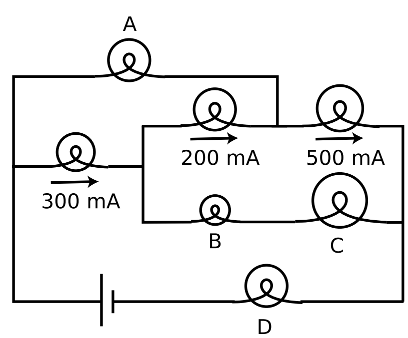
Problems 13 and 14.



Problem 15.



Problem 16.




Problem 17.

# Chapter 5
# Fields of Force

“Okay. Your duties are as follows: Get Breen. I don’t care how you get him, but get him soon. That faker! He posed for twenty years as a scientist without ever being apprehended. Well, I’m going to do some apprehending that’ll make all previous apprehending look like no apprehension at all. You with me?”

“Yes,” said Battle, very much confused. “What’s that thing you have?”

“Piggy-back heat-ray. You transpose the air in its path into an unstable isotope which tends to carry all energy as heat. Then you shoot your juice light, or whatever along the isotopic path and you burn whatever’s on the receiving end. You want a few?”

“No,” said Battle. “I have my gats. What else have you got for offense and defense?” Underbottam opened a cabinet and proudly waved an arm. “Everything,” he said.

“Disintegraters, heat-rays, bombs of every type. And impenetrable shields of energy, massive and portable. What more do I need?”

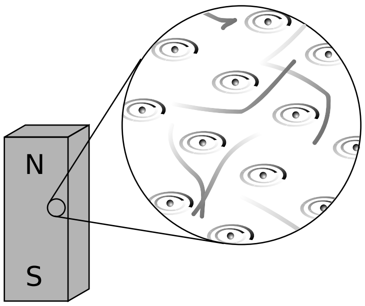
From THE REVERSIBLE REVOLUTIONS by Cecil Corwin, Cosmic Stories, March 1941. Art by Morey, Bok, Kyle, Hunt, Forte. Copyright expired.

Cutting-edge science readily infiltrates popular culture, though sometimes in garbled form. The Newtonian imagination populated the universe mostly with that nice solid stuff called matter, which was made of little hard balls called atoms. In the early twentieth century, consumers of pulp fiction and popularized science began to hear of a new image of the universe, full of x-rays, N-rays, and Hertzian waves. What they were beginning to soak up through their skins was a drastic revision of Newton’s concept of a universe made of chunks of matter which happened to interact via forces. In the newly emerging picture, the universe was *made* of force, or, to be more technically accurate, of ripples in universal fields of force. Unlike the average reader of Cosmic Stories in 1941, you now possess enough technical background to understand what a “force field” really is.

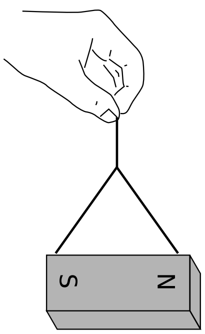
## 5.1 Why fields?

### Time delays in forces exerted at a distance

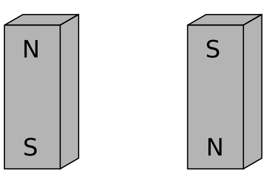
What convinced physicists that they needed this new concept of a field of force? Although we have been dealing mostly with electrical forces, let's start with a magnetic example. (In fact the main reason I've delayed a detailed discussion of magnetism for so long is that mathematical calculations of magnetic effects are handled much more easily with the concept of a field of force.) First a little background leading up to our example. A bar magnet, a, has an axis about which many of the electrons' orbits are oriented. The earth itself is also a magnet, although not a bar-shaped one. The interaction between the earth-magnet and the bar magnet, b, makes them want to line up their axes in opposing directions (in other words such that their electrons rotate in parallel planes, but with one set rotating clockwise and the other counterclockwise as seen looking along the axes). On a smaller scale, any two bar magnets placed near each other will try to align themselves head-to-tail, c.




a / A bar magnet's atoms are (partially) aligned.




b / A bar magnet interacts with our magnetic planet.




c / Magnets aligned north-south.

Now we get to the relevant example. It is clear that two people separated by a paper-thin wall could use a pair of bar magnets to signal to each other. Each person would feel her own magnet trying to twist around in response to any rotation performed by the other person's magnet. The practical range of communication would be very short for this setup, but a sensitive electrical apparatus could pick up magnetic signals from much farther away. In fact, this is not so different from what a radio does: the electrons racing up and down the transmitting antenna create forces on the electrons in the distant receiving antenna. (Both magnetic and electric forces are involved in real radio signals, but we don't need to worry about that yet.)

A question now naturally arises as to whether there is any time delay in this kind of communication via magnetic (and electric) forces. Newton would have thought not, since he conceived of physics in terms of instantaneous action at a distance. We now know, however, that there is such a time delay. If you make a long-distance phone call that is routed through a communications satellite, you should easily be able to detect a delay of about half a second over the signal's round trip of 50,000 miles. Modern measurements have shown that electric, magnetic, and gravitational forces all travel at the speed of light, $3 \times 10^8$ m/s.[1] (In fact, we will soon discuss how light itself is made of electricity and magnetism.)

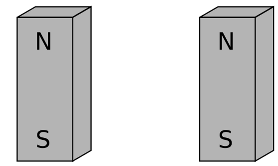
If it takes some time for forces to be transmitted through space, then apparently there is some *thing* that travels *through* space. The fact that the phenomenon travels outward at the same speed in all directions strongly evokes wave metaphors such as ripples on a pond.

### More evidence that fields of force are real: they carry energy.

The smoking-gun argument for this strange notion of traveling force ripples comes from the fact that they carry energy.

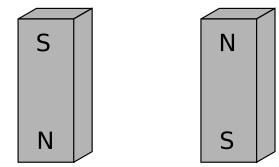
[1] As discussed in book 6 of this series, one consequence of Einstein's theory of relativity is that material objects can never move faster than the speed of light. It can also be shown that signals or information are subject to the same limit.

First suppose that the person holding the bar magnet on the right decides to reverse hers, resulting in configuration d. She had to do mechanical work to twist it, and if she releases the magnet, energy will be released as it flips back to c. She has apparently stored energy by going from c to d. So far everything is easily explained without the concept of a field of force.




d / The second magnet is reversed.

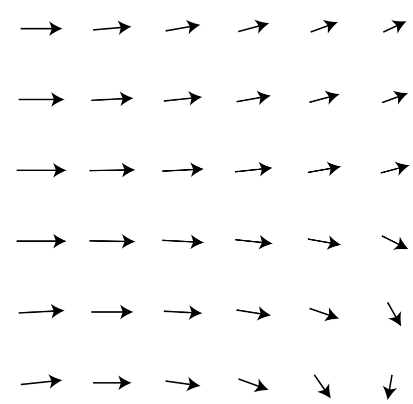
But now imagine that the two people start in position c and then simultaneously flip their magnets extremely quickly to position e, keeping them lined up with each other the whole time. Imagine, for the sake of argument, that they can do this so quickly that each magnet is reversed while the force signal from the other is still in transit. (For a more realistic example, we'd have to have two radio antennas, not two magnets, but the magnets are easier to visualize.) During the flipping, each magnet is still feeling the forces arising from the way the other magnet *used* to be oriented. Even though the two magnets stay aligned during the flip, the time delay causes each person to feel resistance as she twists her magnet around. How can this be? Both of them are apparently doing mechanical work, so they must be storing magnetic energy somehow. But in the traditional Newtonian conception of matter interacting via instantaneous forces at a distance, interaction energy arises from the relative positions of objects that are interacting via forces. If the magnets never changed their orientations relative to each other, how can any magnetic energy have been stored?




e / Both magnets are reversed.

The only possible answer is that the energy must have gone into the magnetic force ripples crisscrossing the space between the magnets. Fields of force apparently carry energy across space, which is strong evidence that they are real things.

This is perhaps not as radical an idea to us as it was to our ancestors. We are used to the idea that a radio transmitting antenna consumes a great deal of power, and somehow spews it out into the universe. A person working around such an antenna needs to be careful not to get too close to it, since all that energy can easily cook flesh (a painful phenomenon known as an "RF burn").



f / The wind patterns in a certain area of the ocean could be charted in a "sea of arrows" representation like this. Each arrow represents both the wind's strength and its direction at a certain location.

## 5.2 The gravitational field

Given that fields of force are real, how do we define, measure, and calculate them? A fruitful metaphor will be the wind patterns experienced by a sailing ship. Wherever the ship goes, it will feel a certain amount of force from the wind, and that force will be in a certain direction. The weather is ever-changing, of course, but for now let's just imagine steady wind patterns. Definitions in physics are operational, i.e., they describe how to measure the thing being defined. The ship's captain can measure the wind's "field of force" by going to the location of interest and determining both the direction of the wind and the strength with which it is blowing. Charting all these measurements on a map leads to a depiction of the field of wind force like the one shown in the figure. This is known as the "sea of arrows" method of visualizing a field.

Now let's see how these concepts are applied to the fundamental force fields of the universe. We'll start with the gravitational field, which is the easiest to understand. As with the wind patterns, we'll start by imagining gravity as a static field, even though the existence of the tides proves that there are continual changes in the gravity field in our region of space. Defining the direction of the gravitational field is easy enough: we simply go to the location of interest and measure the direction of the gravitational force on an object, such as a weight tied to the end of a string.

But how should we define the strength of the gravitational field? Gravitational forces are weaker on the moon than on the earth, but we cannot specify the strength of gravity simply by giving a certain number of newtons. The number of newtons of gravitational force depends not just on the strength of the local gravitational field but also on the mass of the object on which we're testing gravity, our "test mass." A boulder on the moon feels a stronger gravitational force than a pebble on the earth. We can get around this problem by defining the strength of the gravitational field as the force acting on an object, *divided by the object's mass*.

**definition of the gravitational field**
The gravitational field vector, $\mathbf{g}$, at any location in space is found by placing a test mass $m_t$ at that point. The field vector is then given by $\mathbf{g} = \mathbf{F}/m_t$, where $\mathbf{F}$ is the gravitational force on the test mass.

The magnitude of the gravitational field near the surface of the earth is about 9.8 N/kg, and it's no coincidence that this number looks familiar, or that the symbol $\mathbf{g}$ is the same as the one for gravitational acceleration. The force of gravity on a test mass will equal $m_t\mathbf{g}$, where $\mathbf{g}$ is the gravitational acceleration. Dividing by $m_t$ simply gives the gravitational acceleration. Why define a new name and new units for the same old quantity? The main reason is that it prepares us with the right approach for defining other fields.

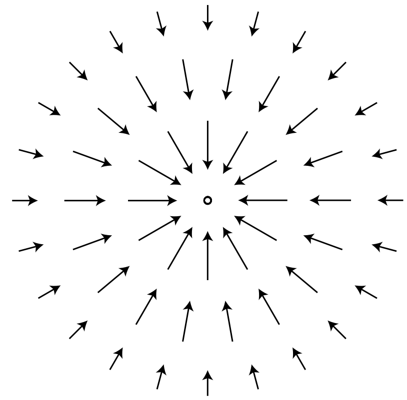
The most subtle point about all this is that the gravitational field tells us about what forces *would* be exerted on a test mass by the earth, sun, moon, and the rest of the universe, *if* we inserted a test mass at the point in question. The field still exists at all the places where we didn't measure it.

*Gravitational field of the earth* *example 1*

▷ What is the magnitude of the earth's gravitational field, in terms of its mass, $M$, and the distance $r$ from its center?

▷ Substituting $|\mathbf{F}| = GMm_t/r^2$ into the definition of the gravitational field, we find $|\mathbf{g}| = GM/r^2$. This expression could be used for the field of any spherically symmetric mass distribution, since the equation we assumed for the gravitational force would apply in any such case.

## Sources and sinks

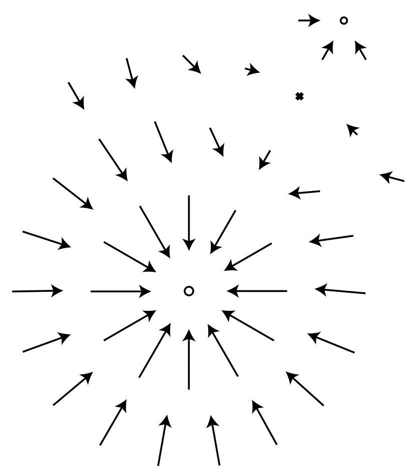


g / The gravitational field surrounding a clump of mass such as the earth.

If we make a sea-of-arrows picture of the gravitational fields surrounding the earth, g, the result is evocative of water going down a drain. For this reason, anything that creates an inward-pointing field around itself is called a sink. The earth is a gravitational sink. The term "source" can refer specifically to things that make outward fields, or it can be used as a more general term for both "outies" and "innies." However confusing the terminology, we know that gravitational fields are only attractive, so we will never find a region of space with an outward-pointing field pattern.

Knowledge of the field is interchangeable with knowledge of its sources (at least in the case of a static, unchanging field). If aliens saw the earth's gravitational field pattern they could immediately infer the existence of the planet, and conversely if they knew the mass of the earth they could predict its influence on the surrounding gravitational field.

## Superposition of fields



h / The gravitational fields of the earth and moon superpose. Note how the fields cancel at one point, and how there is no boundary between the interpenetrating fields surrounding the two bodies.

A very important fact about all fields of force is that when there is more than one source (or sink), the fields add according to the rules of vector addition. The gravitational field certainly will have this property, since it is defined in terms of the force on a test mass, and forces add like vectors. Superposition is an important characteristics of waves, so the superposition property of fields is consistent with the idea that disturbances can propagate outward as waves in a field.

*Reduction in gravity on Io due to Jupiter's gravity* *example 2*

▷ The average gravitational field on Jupiter's moon Io is 1.81 N/kg. By how much is this reduced when Jupiter is directly overhead? Io's orbit has a radius of $4.22\times10^8$ m, and Jupiter's mass is $1.899\times10^{27}$ kg.

▷ By the shell theorem, we can treat the Jupiter as if its mass was all concentrated at its center, and likewise for Io. If we visit Io and land at the point where Jupiter is overhead, we are on the same line as these two centers, so the whole problem can be treated one-dimensionally, and vector addition is just like scalar addition. Let's use positive numbers for downward fields (toward the center of Io) and negative for upward ones. Plugging the appropriate data into the expression derived in example 1, we find that the Jupiter's contribution to the field is $-0.71$ N/kg. Superposition says that we can find the actual gravitational field by adding up the fields created by Io and Jupiter: $1.81-0.71$ N/kg = 1.1 N/kg. You might think that this reduction would create some spectacular effects, and make Io an exciting tourist destination. Actually you would not detect any difference if you flew from one side of Io to the other. This is because your body and Io both experience Jupiter's gravity, so you follow the same orbital curve through the space around Jupiter.

## Gravitational waves

A source that sits still will create a static field pattern, like a steel ball sitting peacefully on a sheet of rubber. A moving source will create a spreading wave pattern in the field, like a bug thrashing on the surface of a pond. Although we have started with the gravitational field as the simplest example of a static field, stars and planets do more stately gliding than thrashing, so gravitational waves are not easy to detect. Newton's theory of gravity does not describe gravitational waves, but they are predicted by Einstein's general theory of relativity. J.H. Taylor and R.A. Hulse were awarded the Nobel Prize in 1993 for giving indirect evidence that Einstein's waves actually exist. They discovered a pair of exotic, ultra-dense stars called neutron stars orbiting one another very closely, and showed that they were losing orbital energy at the rate predicted by Einstein's theory.

i / The part of the LIGO gravity wave detector at Hanford Nuclear Reservation, near Richland, Washington. The other half of the detector is in Louisiana.

A Caltech-MIT collaboration has built a pair of gravitational wave detectors called LIGO to search for more direct evidence of gravitational waves. Since they are essentially the most sensitive vibration detectors ever made, they are located in quiet rural areas, and signals will be compared between them to make sure that they were not due to passing trucks. The project began operating at full sensitivity in 2005, and is now able to detect a vibration that causes a change of $10^{-18}$ m in the distance between the mirrors at the ends of the 4-km vacuum tunnels. This is a thousand times less than the size of an atomic nucleus! There is only enough funding to keep the detectors operating for a few more years, so the physicists can only hope that during that time, somewhere in the universe, a sufficiently violent cataclysm will occur to make a detectable gravitational wave. (More accurately, they want the wave to arrive in our solar system during that time, although it will have been produced millions of years before.)

## 5.3 The electric field

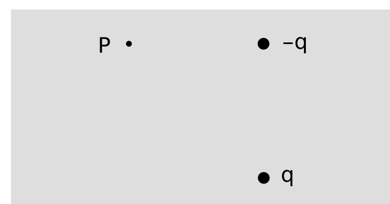
### Definition

The definition of the electric field is directly analogous to, and has the same motivation as, the definition of the gravitational field:

> **definition of the electric field**
> The electric field vector, $\mathbf{E}$, at any location in space is found by placing a test charge $q_t$ at that point. The electric field vector is then given by $\mathbf{E} = \mathbf{F}/q_t$, where $\mathbf{F}$ is the electric force on the test charge.

Charges are what create electric fields. Unlike gravity, which is always attractive, electricity displays both attraction and repulsion. A positive charge is a source of electric fields, and a negative one is a sink.

The most difficult point about the definition of the electric field is that the force on a negative charge is in the opposite direction compared to the field. This follows from the definition, since dividing a vector by a negative number reverses its direction. It's as though we had some objects that fell upward instead of down.




j / Example 3.

*self-check A*

Find an equation for the magnitude of the field of a single point charge $Q$. ▷ Answer, p. 195

*Superposition of electric fields* *example 3*

▷ Charges $q$ and $-q$ are at a distance $b$ from each other, as shown in the figure. What is the electric field at the point P, which lies at a third corner of the square?

▷ The field at P is the vector sum of the fields that would have been created by the two charges independently. Let positive $x$ be to the right and let positive $y$ be up.

Negative charges have fields that point at them, so the charge $-q$ makes a field that points to the right, i.e., has a positive $x$ component. Using the answer to the self-check, we have

$$E_{-q,x} = \frac{kq}{b^2}$$
$$E_{-q,y} = 0 \qquad .$$

Note that if we had blindly ignored the absolute value signs and plugged in $-q$ to the equation, we would have incorrectly concluded that the field went to the left.

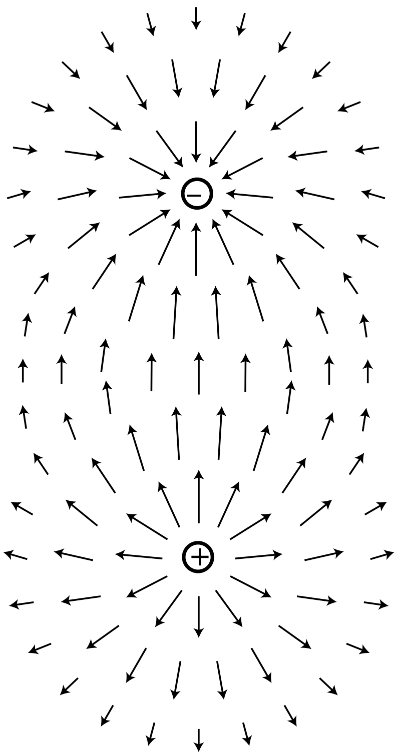
By the Pythagorean theorem, the positive charge is at a distance $\sqrt{2}b$ from P, so the magnitude of its contribution to the field is $E = kq/2b^2$.

Positive charges have fields that point away from them, so the field vector is at an angle of 135 ° counterclockwise from the $x$ axis.

$$
\begin{aligned}
E_{q,x} &= \frac{kq}{2b^2}\cos 135^\circ \\
&= -\frac{kq}{2^{3/2}b^2} \\
E_{q,y} &= \frac{kq}{2b^2}\sin 135^\circ \\
&= \frac{kq}{2^{3/2}b^2}
\end{aligned}
$$

The total field is

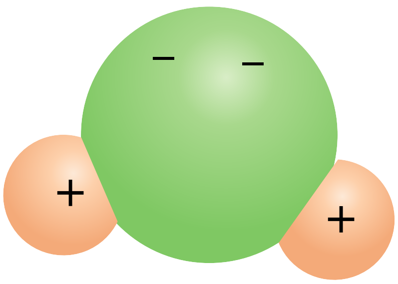
$$
\begin{aligned}
E_x &= \left(1 - 2^{-3/2}\right)\frac{kq}{b^2} \\
E_y &= \frac{kq}{2^{3/2}b^2}
\end{aligned}
$$



k / A dipole field. Electric fields diverge from a positive charge and converge on a negative charge.

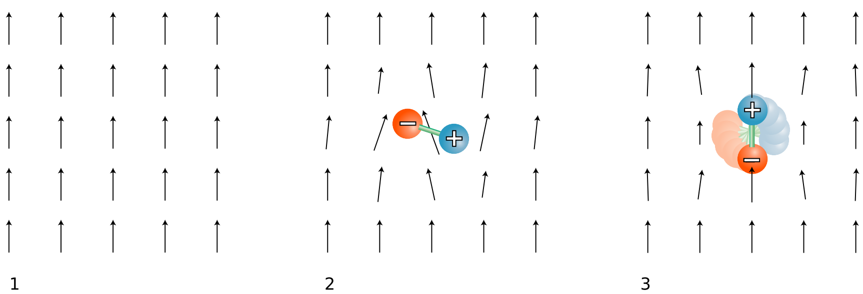
### Dipoles

The simplest set of sources that can occur with electricity but not with gravity is the *dipole*, consisting of a positive charge and a negative charge with equal magnitudes. More generally, an electric dipole can be any object with an imbalance of positive charge on one side and negative on the other. A water molecule, l, is a dipole because the electrons tend to shift away from the hydrogen atoms and onto the oxygen atom.



l / A water molecule is a dipole.

Your microwave oven acts on water molecules with electric fields. Let us imagine what happens if we start with a uniform electric field, m/1, made by some external charges, and then insert a dipole, m/2, consisting of two charges connected by a rigid rod. The dipole disturbs the field pattern, but more important for our present purposes is that it experiences a torque. In this example, the positive charge feels an upward force, but the negative charge is pulled down. The




m / 1. A uniform electric field created by some charges “off-stage.” 2. A dipole is placed in the field. 3. The dipole aligns with the field.

result is that the dipole wants to align itself with the field, m/3. The microwave oven heats food with electrical (and magnetic) waves. The alternation of the torque causes the molecules to wiggle and increase the amount of random motion. The slightly vague definition of a dipole given above can be improved by saying that a dipole is any object that experiences a torque in an electric field.

What determines the torque on a dipole placed in an externally created field? Torque depends on the force, the distance from the axis at which the force is applied, and the angle between the force and the line from the axis to the point of application. Let a dipole consisting of charges $+q$ and $-q$ separated by a distance $\ell$ be placed in an external field of magnitude $|\mathbf{E}|$, at an angle $\theta$ with respect to the field. The total torque on the dipole is

$$\begin{aligned}\tau &= \frac{\ell}{2}q|\mathbf{E}|\sin\theta + \frac{\ell}{2}q|\mathbf{E}|\sin\theta \\ &= \ell q|\mathbf{E}|\sin\theta \qquad .\end{aligned}$$

(Note that even though the two forces are in opposite directions, the torques do not cancel, because they are both trying to twist the dipole in the same direction.) The quantity is called the dipole moment, notated $D$. (More complex dipoles can also be assigned a dipole moment — they are defined as having the same dipole moment as the two-charge dipole that would experience the same torque.)

*Dipole moment of a molecule of NaCl gas* *example 4*

▷ In a molecule of NaCl gas, the center-to-center distance between the two atoms is about 0.6 nm. Assuming that the chlorine completely steals one of the sodium's electrons, compute the magnitude of this molecule's dipole moment.

▷ The total charge is zero, so it doesn't matter where we choose the origin of our coordinate system. For convenience, let's choose it to be at one of the atoms, so that the charge on that atom doesn't contribute to the dipole moment. The magnitude of the dipole moment is then

$$\begin{aligned}D &= (6\times10^{-10}\text{ m})(e) \\ &= (6\times10^{-10}\text{ m})(1.6\times10^{-19}\text{ C}) \\ &= 1\times10^{-28}\text{ C}\cdot\text{m}\end{aligned}$$

### Alternative definition of the electric field

The behavior of a dipole in an externally created field leads us to an alternative definition of the electric field:

**alternative definition of the electric field**
The electric field vector, $E$, at any location in space is defined by observing the torque exerted on a test dipole $D_t$ placed there. The direction of the field is the direction in which the field tends to align a dipole (from $-$ to $+$), and the field's magnitude is $|\mathbf{E}| = \tau/D_t\sin\theta$.

The main reason for introducing a second definition for the same concept is that the magnetic field is most easily defined using a similar approach.

### Voltage related to electric field

Voltage is potential energy per unit charge, and electric field is force per unit charge. We can therefore relate voltage and field if we start from the relationship between potential energy and force,

$$\Delta PE = -Fd \quad , \qquad \text{[assuming constant force and motion parallel to the force]}$$

and divide by charge,

$$\Delta PE = -Fd \quad , \qquad \text{[assuming constant force and motion parallel to the force]}$$

giving

$$\Delta V = -Ed \quad , \qquad \text{[assuming constant force and motion parallel to the force]}$$

In other words, the difference in voltage between two points equals the electric field strength multiplied by the distance between them. The interpretation is that a strong electric field is a region of space where the voltage is rapidly changing. By analogy, a steep hillside is a place on the map where the altitude is rapidly changing.

*Field generated by an electric eel* — *example 5*

▷ Suppose an electric eel is 1 m long, and generates a voltage difference of 1000 volts between its head and tail. What is the electric field in the water around it?

▷ We are only calculating the amount of field, not its direction, so we ignore positive and negative signs. Subject to the possibly inaccurate assumption of a constant field parallel to the eel's body, we have

$$|\mathbf{E}| = \frac{\Delta V}{\Delta x} = 1000\ \text{V/m} \quad .$$

*Relating the units of electric field and voltage* — *example 6*

From our original definition of the electric field, we expect it to have units of newtons per coulomb, N/C. The example above, however, came out in volts per meter, V/m. Are these inconsistent? Let's reassure ourselves that this all works. In this kind of situation, the best strategy is usually to simplify the more complex units so that they involve only mks units and coulombs. Since voltage is defined as electrical energy per unit charge, it has units of J/C:

$$\frac{\text{V}}{\text{m}} = \frac{\text{J/C}}{\text{m}} = \frac{\text{J}}{\text{C}\cdot\text{m}} \quad .$$

To connect joules to newtons, we recall that work equals force times distance, so J = N · m, so

$$\begin{aligned}\frac{\text{V}}{\text{m}} &= \frac{\text{N}\cdot\text{m}}{\text{C}\cdot\text{m}} \\ &= \frac{\text{N}}{\text{C}}\end{aligned}$$

As with other such difficulties with electrical units, one quickly begins to recognize frequently occurring combinations.

## Discussion Questions

**A** In the definition of the electric field, does the test charge need to be 1 coulomb? Does it need to be positive?

**B** Does a charged particle such as an electron or proton feel a force from its own electric field?

**C** Is there an electric field surrounding a wall socket that has nothing plugged into it, or a battery that is just sitting on a table?

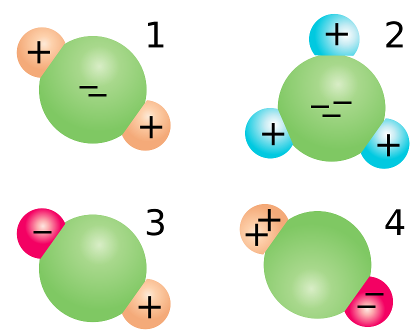
**D** In a flashlight powered by a battery, which way do the electric fields point? What would the fields be like inside the wires? Inside the filament of the bulb?

**E** Criticize the following statement: "An electric field can be represented by a sea of arrows showing how current is flowing."

**F** The field of a point charge, $|\mathbf{E}| = kQ/r^2$, was derived in the self-check above. How would the field pattern of a uniformly charged sphere compare with the field of a point charge?

**G** The interior of a perfect electrical conductor in equilibrium must have zero electric field, since otherwise the free charges within it would be drifting in response to the field, and it would not be in equilibrium. What about the field right at the surface of a perfect conductor? Consider the possibility of a field perpendicular to the surface or parallel to it.

**H** Compare the dipole moments of the molecules and molecular ions shown in the figure.



n / Discussion question H.

**I** Small pieces of paper that have not been electrically prepared in any way can be picked up with a charged object such as a charged piece of tape. In our new terminology, we could describe the tape's charge as inducing a dipole moment in the paper. Can a similar technique be used to induce not just a dipole moment but a charge?

**J** The earth and moon are fairly uneven in size and far apart, like a baseball and a ping-pong ball held in your outstretched arms. Imagine instead a planetary system with the character of a double planet: two planets of equal size, close together. Sketch a sea of arrows diagram of their gravitational field.

## 5.4 $\int$ Voltage for Nonuniform Fields

The calculus-savvy reader will have no difficulty generalizing the field-voltage relationship to the case of a varying field. The potential energy associated with a varying force is

$$\Delta PE = -\int F\,\mathrm{d}x \qquad , \qquad \text{[one dimension]}$$

so for electric fields we divide by $q$ to find

$$\Delta V = -\int E\,\mathrm{d}x \qquad , \qquad \text{[one dimension]}$$

Applying the fundamental theorem of calculus yields

$$E = -\frac{\mathrm{d}V}{\mathrm{d}x} \qquad . \qquad \text{[one dimension]}$$

*Voltage associated with a point charge* *example 7*

▷ What is the voltage associated with a point charge?

▷ As derived previously in self-check A on page 129, the field is

$$|\mathbf{E}| = \frac{kQ}{r^2}$$

The difference in voltage between two points on the same radius line is

$$\begin{aligned}\Delta V &= -\int \mathrm{d}V \\ &= -\int E_x \mathrm{d}x\end{aligned}$$

In the general discussion above, $x$ was just a generic name for distance traveled along the line from one point to the other, so in this case $x$ really means $r$.

$$\begin{aligned}\Delta V &= -\int_{r_1}^{r_2} E_r \mathrm{d}r \\ &= -\int_{r_1}^{r_2} \frac{kQ}{r^2}\mathrm{d}r \\ &= \frac{kQ}{r}\Big]_{r_1}^{r_2} \\ &= \frac{kQ}{r_2} - \frac{kQ}{r_1} \qquad .\end{aligned}$$

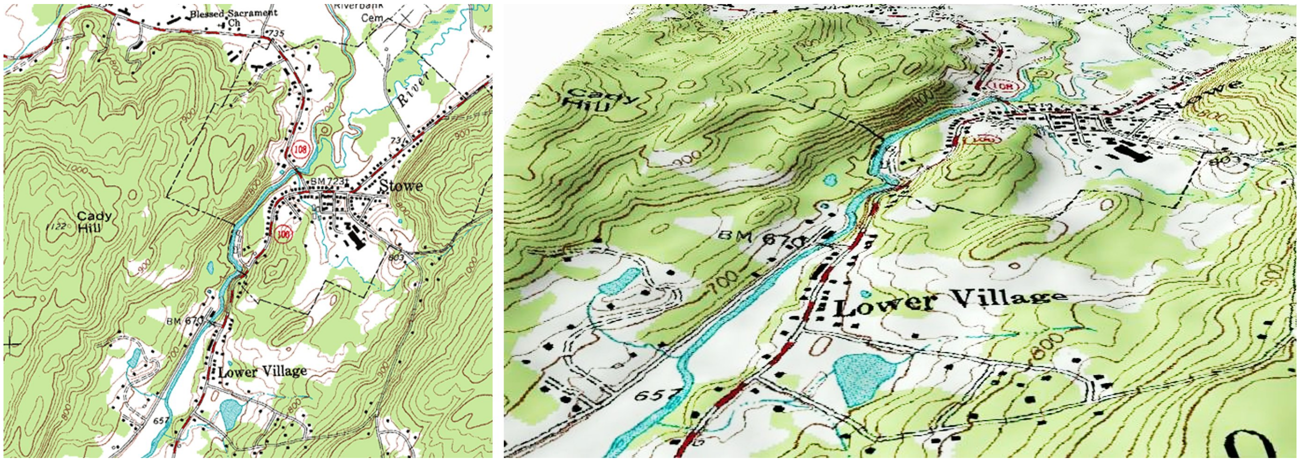
The standard convention is to use $r_1 = \infty$ as a reference point, so that the voltage at any distance $r$ from the charge is

$$V = \frac{kQ}{r} \qquad .$$

The interpretation is that if you bring a positive test charge closer to a positive charge, its electrical energy is increased; if it was released, it would spring away, releasing this as kinetic energy.

*self-check B*

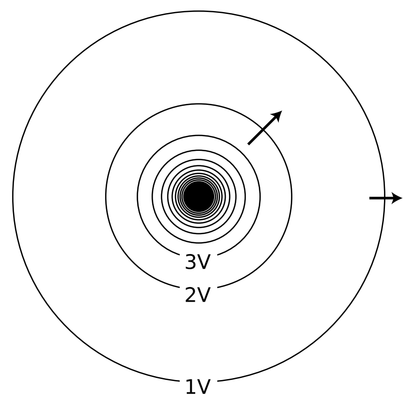
Show that you can recover the expression for the field of a point charge by evaluating the derivative $E_x = -\mathrm{d}V/\mathrm{d}x$. ▷ Answer, p. 195




o / Left: A topographical map of Stowe, Vermont. From one constant-height line to the next is a height difference of 200 feet. Lines far apart, as in the lower village, indicate relatively flat terrain, while lines close together, like the ones to the west of the main town, represent a steep slope. Streams flow downhill, perpendicular to the constant-height lines. Right: The same map has been redrawn in perspective, with shading to suggest relief.

## 5.5 Two or Three Dimensions

The topographical map shown in figure o suggests a good way to visualize the relationship between field and voltage in two dimensions. Each contour on the map is a line of constant height; some of these are labeled with their elevations in units of feet. Height is related to gravitational potential energy, so in a gravitational analogy, we can think of height as representing voltage. Where the contour lines are far apart, as in the town, the slope is gentle. Lines close together indicate a steep slope.

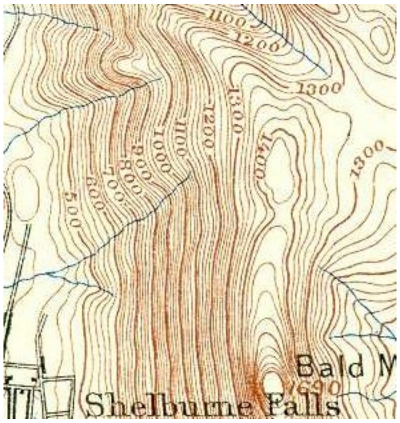



p / The constant-voltage curves surrounding a point charge. Near the charge, the curves are so closely spaced that they blend together on this drawing due to the finite width with which they were drawn. Some electric fields are shown as arrows.

If we walk along a straight line, say straight east from the town, then height (voltage) is a function of the east-west coordinate $x$. Using the usual mathematical definition of the slope, and writing $V$ for the height in order to remind us of the electrical analogy, the slope along such a line is $\Delta V/\Delta \mathrm{x}$. If the slope isn't constant, we either need to use the slope of the $V-x$ graph, or use calculus and talk about the derivative $\mathrm{d}V/\mathrm{d}x$.

What if everything isn't confined to a straight line? Water flows downhill. Notice how the streams on the map cut perpendicularly through the lines of constant height.

It is possible to map voltages in the same way, as shown in figure p. The electric field is strongest where the constant-voltage curves are closest together, and the electric field vectors always point perpendicular to the constant-voltage curves.




q / Self-check C.

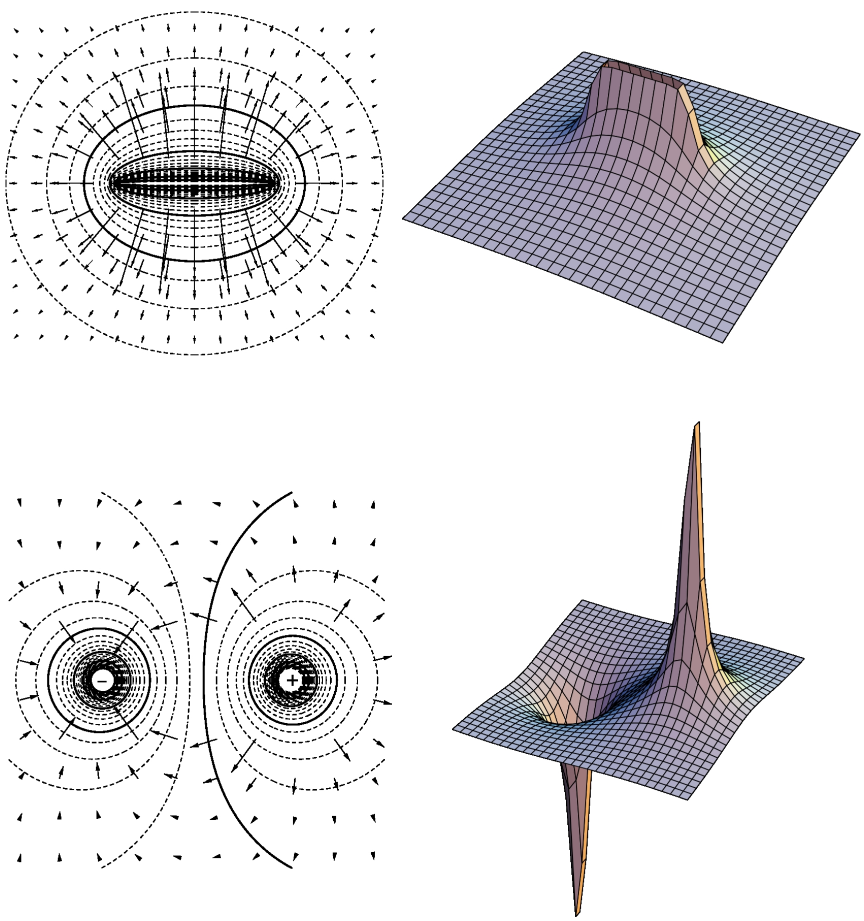
Figure r shows some examples of ways to visualize field and voltage patterns.

Mathematically, the calculus of section 5.4 generalizes to three dimensions as follows:

$$E_x = -\mathrm{d}V/\mathrm{d}x$$
$$E_y = -\mathrm{d}V/\mathrm{d}y$$
$$E_z = -\mathrm{d}V/\mathrm{d}z$$

*self-check C*

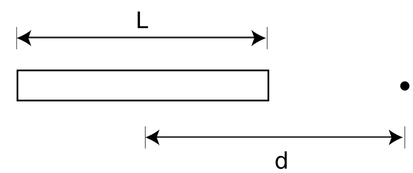
Imagine that the topographical map in figure q represents voltage rather than height. (a) Consider the stream the starts near the center of the map. Determine the positive and negative signs of $dV/dx$ and $dV/dy$, and relate these to the direction of the force that is pushing the current forward against the resistance of friction. (b) If you wanted to find a lot of electric charge on this map, where would you look? ▷ Answer, p. 196



r / Two-dimensional field and voltage patterns. Top: A uniformly charged rod. Bottom: A dipole. In each case, the diagram on the left shows the field vectors and constant-voltage curves, while the one on the right shows the voltage (up-down coordinate) as a function of x and y. Interpreting the field diagrams: Each arrow represents the field at the point where its tail has been positioned. For clarity, some of the arrows in regions of very strong field strength are not shown Î¡ they would be too long to show. Interpreting the constant-voltage curves: In regions of very strong fields, the curves are not shown because they would merge together to make solid black regions. Interpreting the perspective plots: Keep in mind that even though we're visualizing things in three dimensions, these are really two-dimensional voltage patterns being represented. The third (up-down) dimension represents voltage, not position.

## 5.6 $\int\star$ Electric Field of a Continuous Charge Distribution

Charge really comes in discrete chunks, but often it is mathematically convenient to treat a set of charges as if they were like a continuous fluid spread throughout a region of space. For example, a charged metal ball will have charge spread nearly uniformly all over its surface, and in for most purposes it will make sense to ignore the fact that this uniformity is broken at the atomic level. The electric field made by such a continuous charge distribution is the sum of the fields created by every part of it. If we let the "parts" become infinitesimally small, we have a sum of an infinite number of infinitesimal numbers, which is an integral. If it was a discrete sum, we would have a total electric field in the $x$ direction that was the sum of all the $x$ components of the individual fields, and similarly we'd have sums for the $y$ and $z$ components. In the continuous case, we have three integrals.




s / Example 8.

*Field of a uniformly charged rod* *example 8*

▷ A rod of length $L$ has charge $Q$ spread uniformly along it. Find the electric field at a point a distance $d$ from the center of the rod, along the rod's axis.

▷ This is a one-dimensional situation, so we really only need to do a single integral representing the total field along the axis. We imagine breaking the rod down into short pieces of length $\mathrm{d}z$, each with charge $\mathrm{d}q$. Since charge is uniformly spread along the rod, we have $\mathrm{d}q = \lambda\mathrm{d}z$, where $\lambda = Q/L$ (Greek lambda) is the charge per unit length, in units of coulombs per meter. Since the pieces are infinitesimally short, we can treat them as point charges and use the expression $k\mathrm{d}q/r^2$ for their contributions to the field, where $r = d - z$ is the distance from the charge at $z$ to the point in which we are interested.

$$\begin{aligned} E_z &= \int \frac{k\mathrm{d}q}{r^2} \\ &= \int_{-L/2}^{+L/2} \frac{k\lambda\mathrm{d}z}{r^2} \\ &= k\lambda \int_{-L/2}^{+L/2} \frac{\mathrm{d}z}{(d-z)^2} \end{aligned}$$

The integral can be looked up in a table, or reduced to an elementary form by substituting a new variable for $d - z$. The result is

$$\begin{aligned} E_z &= k\lambda \left(\frac{1}{d-z}\right)_{-L/2}^{+L/2} \\ &= \frac{kQ}{L}\left(\frac{1}{d-L/2} - \frac{1}{d+L/2}\right) \quad . \end{aligned}$$

For large values of $d$, this expression gets smaller for two reasons: (1) the denominators of the fractions become large, and (2) the two fractions become nearly the same, and tend to cancel out. This makes sense, since the field should get weaker as we get farther away from the charge. In fact, the field at large distances must approach $kQ/d^2$, since from a great distance, the rod looks like a point.

It's also interesting to note that the field becomes infinite at the ends of the rod, but is not infinite on the interior of the rod. Can you explain physically why this happens?

## Summary

### Selected Vocabulary

| | |
|---|---|
| field | a property of a point in space describing the forces that would be exerted on a particle if it was there |
| sink-SHARED | a point at which field vectors converge |
| source | a point from which field vectors diverge; often used more inclusively to refer to points of either convergence or divergence |
| electric field | the force per unit charge exerted on a test charge at a given point in space |
| gravitational field | the force per unit mass exerted on a test mass at a given point in space |
| electric dipole | an object that has an imbalance between positive charge on one side and negative charge on the other; an object that will experience a torque in an electric field |

### Notation

| | |
|---|---|
| $\mathbf{g}$ | the gravitational field |
| $\mathbf{E}$ | the electric field |
| $D$ | an electric dipole moment |

### Other Terminology and Notation

| | |
|---|---|
| $d$, $p$, $m$ | other notations for the electric dipole moment |

### Summary

Newton conceived of a universe where forces reached across space instantaneously, but we now know that there is a delay in time before a change in the configuration of mass and charge in one corner of the universe will make itself felt as a change in the forces experienced far away. We imagine the outward spread of such a change as a ripple in an invisible universe-filling *field of force.*

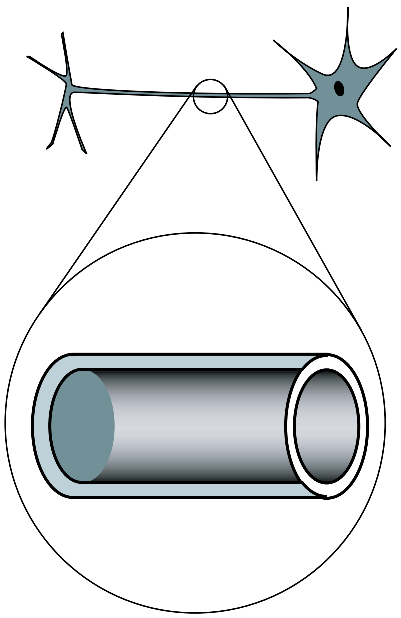
We define the *gravitational field* at a given point as the force per unit mass exerted on objects inserted at that point, and likewise the *electric field* is defined as the force per unit charge. These fields are vectors, and the fields generated by multiple sources add according to the rules of vector addition.

When the electric field is constant, the voltage difference between two points lying on a line parallel to the field is related to the field by the equation $\Delta V = —Ed$, where $d$ is the distance between the two points.

## Problems

**Key**

- √ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ⋆ A difficult problem.



Problem 1.

**1** In our by-now-familiar neuron, the voltage difference between the inner and outer surfaces of the cell membrane is about $V_{out} - V_{in} = -70$ mV in the resting state, and the thickness of the membrane is about 6.0 nm (i.e. only about a hundred atoms thick). What is the electric field inside the membrane? √

**2** The gap between the electrodes in an automobile engine's spark plug is 0.060 cm. To produce an electric spark in a gasoline-air mixture, an electric field of $3.0\times10^6$ V/m must be achieved. On starting a car, what minimum voltage must be supplied by the ignition circuit? Assume the field is uniform. √
(b) The small size of the gap between the electrodes is inconvenient because it can get blocked easily, and special tools are needed to measure it. Why don't they design spark plugs with a wider gap?

**3** (a) At time $t = 0$, a positively charged particle is placed, at rest, in a vacuum, in which there is a uniform electric field of magnitude $E$. Write an equation giving the particle's speed, $v$, in terms of $t$, $E$, and its mass and charge $m$ and $q$. √
(b) If this is done with two different objects and they are observed to have the same motion, what can you conclude about their masses and charges? (For instance, when radioactivity was discovered, it was found that one form of it had the same motion as an electron in this type of experiment.)

**4** Show that the magnitude of the electric field produced by a simple two-charge dipole, at a distant point along the dipole's axis, is to a good approximation proportional to $D/r^3$, where $r$ is the distance from the dipole. [Hint: Use the approximation $(1+\epsilon)^p \approx 1+p\epsilon$, which is valid for small $\epsilon$.] ⋆

**5** Given that the field of a dipole is proportional to $D/r^3$ (see previous problem), show that its voltage varies as $D/r^2$. (Ignore positive and negative signs and numerical constants of proportionality.) ∫

**6** A carbon dioxide molecule is structured like O-C-O, with all three atoms along a line. The oxygen atoms grab a little bit of extra negative charge, leaving the carbon positive. The molecule's symmetry, however, means that it has no overall dipole moment, unlike a V-shaped water molecule, for instance. Whereas the voltage of a dipole of magnitude $D$ is proportional to $D/r^2$ (problem 5), it turns out that the voltage of a carbon dioxide molecule along its axis equals $k/r^3$, where $r$ is the distance from the molecule and $k$

is a constant. What would be the electric field of a carbon dioxide molecule at a distance $r$? $\int$

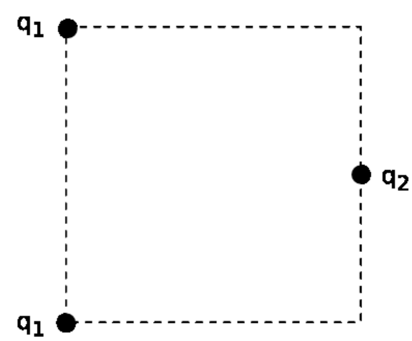
**7** A proton is in a region in which the electric field is given by $E = a + bx^3$. If the proton starts at rest at $x_1 = 0$, find its speed, $v$, when it reaches position $x_2$. Give your answer in terms of $a, b$, $x_2$, and $e$ and $m$, the charge and mass of the proton. $\checkmark$ $\int$

**8** Consider the electric field created by a uniform ring of total charge $q$ and radius $b$. (a) Show that the field at a point on the ring's axis at a distance $a$ from the plane of the ring is $kqa(a^2 + b^2)^{-3/2}$. (b) Show that this expression has the right behavior for $a = 0$ and for $a$ much greater than $b$. $\star$

**9** Consider the electric field created by an infinite uniformly charged plane. Starting from the result of problem 8, show that the field at any point is $2\pi k\sigma$, where $\sigma$ is the density of charge on the plane, in units of coulombs per square meter. Note that the result is independent of the distance from the plane. [Hint: Slice the plane into infinitesimal concentric rings, centered at the point in the plane closest to the point at which the field is being evaluated. Integrate the rings' contributions to the field at this point to find the total field.]
▷ Solution, p. 197 $\int$

**10** Consider the electric field created by a uniformly charged cylinder that extends to infinity in one direction. (a) Starting from the result of problem 8, show that the field at the center of the cylinder's mouth is $2\pi k\sigma$, where $\sigma$ is the density of charge on the cylinder, in units of coulombs per square meter. [Hint: You can use a method similar to the one in problem 9.] (b) This expression is independent of the radius of the cylinder. Explain why this should be so. For example, what would happen if you doubled the cylinder's radius? $\int$

**11** Three charges are arranged on a square as shown. All three charges are positive. What value of $q_2/q_1$ will produce zero electric field at the center of the square? ▷ Solution, p. 197



Problem 11.

a / The first two humans to know what starlight was: James Clerk Maxwell and Katherine Maxwell, 1869.

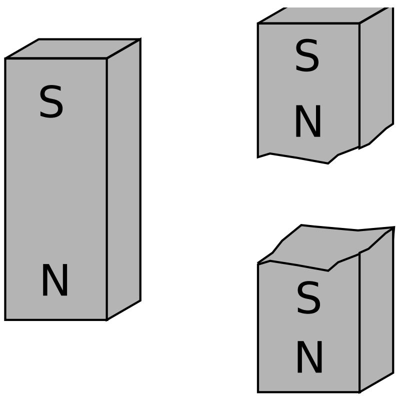
# Chapter 6

# Electromagnetism

In this chapter we discuss the intimate relationship between magnetism and electricity discovered by James Clerk Maxwell. Maxwell realized that light was a wave made up of electric and magnetic fields linked to each other. He is said to have gone for a walk with his wife one night and told her that she was the only other person in the world who knew what starlight really was.

## 6.1 The Magnetic Field

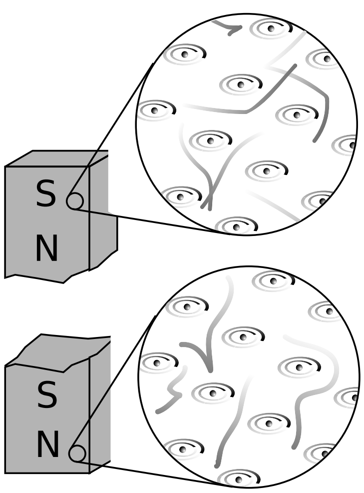
### No magnetic monopoles



b / Breaking a bar magnet in half doesn't create two monopoles, it creates two smaller dipoles.

If you could play with a handful of electric dipoles and a handful of bar magnets, they would appear very similar. For instance, a pair of bar magnets wants to align themselves head-to-tail, and a pair of electric dipoles does the same thing. (It is unfortunately not that easy to make a permanent electric dipole that can be handled like this, since the charge tends to leak.)

You would eventually notice an important difference between the two types of objects, however. The electric dipoles can be broken apart to form isolated positive charges and negative charges. The two-ended device can be broken into parts that are not two-ended. But if you break a bar magnet in half, b, you will find that you have simply made two smaller two-ended objects.



c / An explanation at the atomic level.

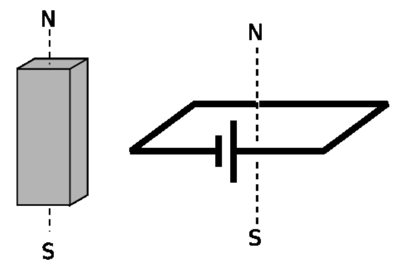
The reason for this behavior is not hard to divine from our microscopic picture of permanent iron magnets. An electric dipole has extra positive "stuff" concentrated in one end and extra negative in the other. The bar magnet, on the other hand, gets its magnetic properties not from an imbalance of magnetic "stuff" at the two ends but from the orientation of the rotation of its electrons. One end is the one from which we could look down the axis and see the electrons rotating clockwise, and the other is the one from which they would appear to go counterclockwise. There is no difference between the "stuff" in one end of the magnet and the other, c.

Nobody has ever succeeded in isolating a single magnetic pole. In technical language, we say that magnetic monopoles do not seem to exist. Electric monopoles *do* exist — that's what charges are.

Electric and magnetic forces seem similar in many ways. Both act at a distance, both can be either attractive or repulsive, and both are intimately related to the property of matter called charge. (Recall that magnetism is an interaction between moving charges.) Physicists's aesthetic senses have been offended for a long time because this seeming symmetry is broken by the existence of electric monopoles and the absence of magnetic ones. Perhaps some exotic form of matter exists, composed of particles that are magnetic monopoles. If such particles could be found in cosmic rays or moon rocks, it would be evidence that the apparent asymmetry was only an asymmetry in the composition of the universe, not in the laws of physics. For these admittedly subjective reasons, there have been several searches for magnetic monopoles. Experiments have been performed, with negative results, to look for magnetic monopoles embedded in ordinary matter. Soviet physicists in the 1960$s$ made exciting claims that they had created and detected magnetic monopoles in particle accelerators, but there was no success in attempts to reproduce the results there or at other accelerators. The most recent search for magnetic monopoles, done by reanalyzing data from the search for the top quark at Fermilab, turned up no candidates, which shows that either monopoles don't exist in nature or they are extremely massive and thus hard to create in accelerators.

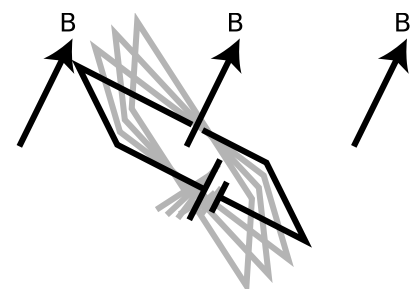
### Definition of the magnetic field

Since magnetic monopoles don't seem to exist, it would not make much sense to define a magnetic field in terms of the force on a test monopole. Instead, we follow the philosophy of the alternative definition of the electric field, and define the field in terms of the torque on a magnetic test dipole. This is exactly what a magnetic compass does: the needle is a little iron magnet which acts like a magnetic dipole and shows us the direction of the earth's magnetic field.




d / A standard dipole made from a square loop of wire shorting across a battery. It acts very much like a bar magnet, but its strength is more easily quantified.

To define the strength of a magnetic field, however, we need some way of defining the strength of a test dipole, i.e., we need a definition of the magnetic dipole moment. We could use an iron permanent magnet constructed according to certain specifications, but such an object is really an extremely complex system consisting of many iron atoms, only some of which are aligned. A more fundamental standard dipole is a square current loop. This could be little resistive circuit consisting of a square of wire shorting across a battery.




e / A dipole tends to align itself to the surrounding magnetic field.

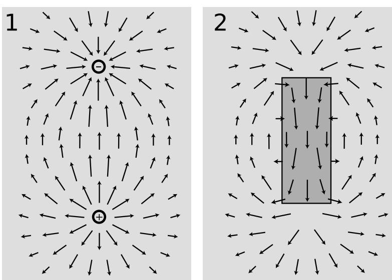
We will find that such a loop, when placed in a magnetic field, experiences a torque that tends to align plane so that its face points in a certain direction. (Since the loop is symmetric, it doesn't care if we rotate it like a wheel without changing the plane in which it lies.) It is this preferred facing direction that we will end up defining as the direction of the magnetic field.

Experiments show if the loop is out of alignment with the field, the torque on it is proportional to the amount of current, and also to the interior area of the loop. The proportionality to current makes sense, since magnetic forces are interactions between moving charges, and current is a measure of the motion of charge. The proportionality to the loop's area is also not hard to understand, because increasing the length of the sides of the square increases both the amount of charge contained in this circular "river" and the amount of leverage supplied for making torque. Two separate physical reasons for a proportionality to length result in an overall proportionality to length squared, which is the same as the area of the loop. For these reasons, we define the magnetic dipole moment of a square current loop as

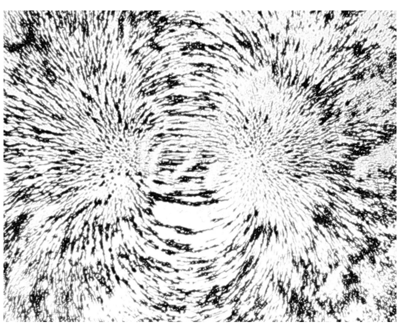
$$D_m = IA \qquad , \qquad \text{[definition of the magnetic dipole moment of a square current loop]}$$

We now define the magnetic field in a manner entirely analogous to the second definition of the electric field:

**definition of the magnetic field**
The magnetic field vector, $\mathbf{B}$, at any location in space is defined by observing the torque exerted on a magnetic test dipole $D_{mt}$ consisting of a square current loop. The field's magnitude is $|\mathbf{B}| = \tau/D_{mt}\sin\theta$, where $\theta$ is the angle by which the loop is misaligned. The direction of the field is perpendicular to the loop; of the two perpendiculars, we choose the one such that if we look along it, the loop's current is counterclockwise.




f / Electric fields, 1, have sources and sinks, but magnetic fields, 2, don't.



g / The magnetic field pattern of a bar magnet. This picture was made by putting iron filings on a piece of paper, and bringing a bar magnet up underneath it. Note how the field pattern passes across the body of the magnet, forming closed loops, as in figure f/2. There are no sources or sinks.

We find from this definition that the magnetic field has units of $\text{N}\cdot\text{m}/\text{A}\cdot\text{m}^2 = \text{N}/\text{A}\cdot\text{m}$. This unwieldy combination of units is abbreviated as the tesla, 1 T = 1 N/A·m. Refrain from memorizing the part about the counterclockwise direction at the end; in section 6.4 we'll see how to understand this in terms of more basic principles.

The nonexistence of magnetic monopoles means that unlike an electric field, f/1, a magnetic one, f/2, can never have sources or sinks. The magnetic field vectors lead in paths that loop back on themselves, without ever converging or diverging at a point.

## 6.2 Calculating Magnetic Fields and Forces

### Magnetostatics

Our study of the electric field built on our previous understanding of electric forces, which was ultimately based on Coulomb's law for the electric force between two point charges. Since magnetism is ultimately an interaction between currents, i.e., between moving charges, it is reasonable to wish for a magnetic analog of Coulomb's law, an equation that would tell us the magnetic force between any two moving point charges.

Such a law, unfortunately, does not exist. Coulomb's law describes the special case of electrostatics: if a set of charges is sitting around and not moving, it tells us the interactions among them. Coulomb's law fails if the charges are in motion, since it does not incorporate any allowance for the time delay in the outward propagation of a change in the locations of the charges.

A pair of moving point charges will certainly exert magnetic forces on one another, but their magnetic fields are like the v-shaped bow waves left by boats. Each point charge experiences a magnetic field that originated from the other charge when it was at some previous position. There is no way to construct a force law that tells us the force between them based only on their current positions in space.

There is, however, a science of magnetostatics that covers a great many important cases. Magnetostatics describes magnetic forces among currents in the special case where the currents are steady and continuous, leading to magnetic fields throughout space that do not change over time.

If we cannot build a magnetostatics from a force law for point charges, then where do we start? It can be done, but the level of mathematics required (vector calculus) is inappropriate for this course. Luckily there is an alternative that is more within our reach. Physicists of generations past have used the fancy math to derive simple equations for the fields created by various static current distributions, such as a coil of wire, a circular loop, or a straight wire. Virtually all practical situations can be treated either directly using these equations or by doing vector addition, e.g., for a case like the field of two circular loops whose fields add onto one another. Figure h shows the equations for some of the more commonly encountered configurations, with illustrations of their field patterns.

*Field created by a long, straight wire carrying current I:*

$$B = \frac{\mu_o I}{2\pi r}$$

Here $r$ is the distance from the center of the wire. The field vectors trace circles in planes perpendicular to the wire, going clockwise when viewed from along the direction of the current.

*Field created by a single circular loop of current:*
The field vectors form a dipole-like pattern, coming through the loop and back around on the outside. Each oval path traced out by the field vectors appears clockwise if viewed from along the direction the current is going when it punches through it. There is no simple equation for a field at an arbitrary point in space, but for a point lying *along the central axis* perpendicular to the loop, the field is

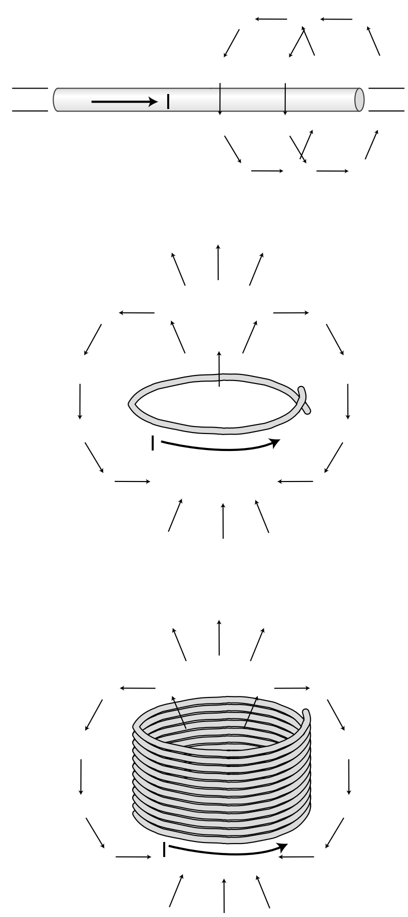
$$B = \frac{1}{2}\mu_o I b^2 \left(b^2 + z^2\right)^{-3/2} \qquad ,$$

where $b$ is the radius of the loop and $z$ is the distance of the point from the plane of the loop.

*Field created by a solenoid (cylindrical coil):*
The field pattern is similar to that of a single loop, but for a long solenoid the paths of the field vectors become very straight on the inside of the coil and on the outside immediately next to the coil. For a sufficiently long solenoid, the interior field also becomes very nearly uniform, with a magnitude of

$$B = \mu_o I N/\ell \qquad ,$$

where $N$ is the number of turns of wire and $\ell$ is the length of the solenoid. The field near the mouths or outside the coil is not constant, and is more difficult to calculate. For a long solenoid, the exterior field is much smaller than the interior field.



h / Some magnetic fields.

Don't memorize the equations! The symbol $\mu_o$ is an abbreviation for the constant $4\pi \times 10^{-7}$ T·m/A. It is the magnetic counterpart of the Coulomb force constant $k$. The Coulomb constant tells us how much electric field is produced by a given amount of charge, while $\mu_o$ relates currents to magnetic fields. Unlike $k$, $\mu_o$ has a definite numerical value because of the design of the metric system.

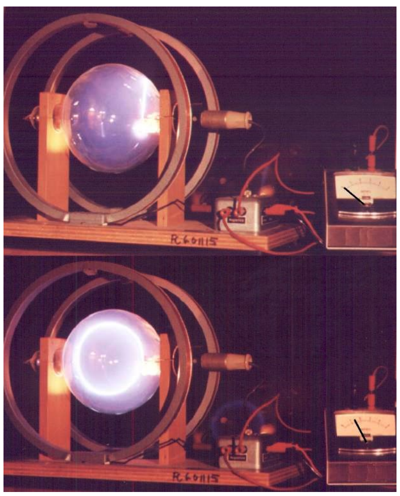
### Force on a charge moving through a magnetic field

We now know how to calculate magnetic fields in some typical situations, but one might also like to be able to calculate magnetic forces, such as the force of a solenoid on a moving charged particle, or the force between two parallel current-carrying wires.

We will restrict ourselves to the case of the force on a charged particle moving through a magnetic field, which allows us to calculate the force between two objects when one is a moving charged particle and the other is one whose magnetic field we know how to find. An example is the use of solenoids inside a TV tube to guide the electron beam as it paints a picture.

Experiments show that the magnetic force on a moving charged particle has a magnitude given by

$$|\mathbf{F}| = q|\mathbf{v}||\mathbf{B}|\sin\theta \qquad ,$$



i / Magnetic forces cause a beam of electrons to move in a circle. The beam is created in a vacuum tube, in which a small amount of hydrogen gas has been left. A few of the electrons strike hydrogen molecules, creating light and letting us see the beam. A magnetic field is produced by passing a current (meter) through the circular coils of wire in front of and behind the tube. In the bottom figure, with the magnetic field turned on, the force perpendicular to the electrons' direction of motion causes them to move in a circle.

where $\mathbf{v}$ is the velocity vector of the particle, and $\theta$ is the angle between the $\mathbf{v}$ and $\mathbf{B}$ vectors. Unlike electric and gravitational forces, magnetic forces do not lie along the same line as the field vector. The force is always *perpendicular* to both $\mathbf{v}$ and $\mathbf{B}$. Given two vectors, there is only one line perpendicular to both of them, so the force vector points in one of the two possible directions along this line. For a positively charged particle, the direction of the force vector can be found as follows. First, position the $\mathbf{v}$ and $\mathbf{B}$ vectors with their tails together. The direction of $\mathbf{F}$ is such that if you sight along it, the $\mathbf{B}$ vector is clockwise from the $\mathbf{v}$ vector; for a negatively charged particle the direction of the force is reversed. Note that since the force is perpendicular to the particle's motion, the magnetic field never does work on it.

*Hallucinations during an MRI scan* *example 1*
During an MRI scan of the head, the patient's nervous system is exposed to intense magnetic fields. The average velocities of the charge-carrying ions in the nerve cells is fairly low, but if the patient moves her head suddenly, the velocity can be high enough that the magnetic field makes significant forces on the ions. This can result in visual and auditory hallucinations, e.g., frying bacon sounds.

## 6.3 Induction

### Electromagnetism and relative motion

The theory of electric and magnetic fields constructed up to this point contains a paradox. One of the most basic principles of physics, dating back to Newton and Galileo and still going strong today, states that motion is relative, not absolute. Thus the laws of physics should not function any differently in a moving frame of reference, or else we would be able to tell which frame of reference was the one in an absolute state of rest. As an example from mechanics, imagine that a child is tossing a ball up and down in the back seat of a moving car. In the child's frame of reference, the car is at rest and the landscape is moving by; in this frame, the ball goes straight up and down, and obeys Newton's laws of motion and Newton's law of gravity. In the frame of reference of an observer watching from the sidewalk, the car is moving and the sidewalk isn't. In this frame, the ball follows a parabolic arc, but it still obeys Newton's laws.

j / Michael Faraday (1791-1867), the son of a poor blacksmith, discovered induction experimentally.

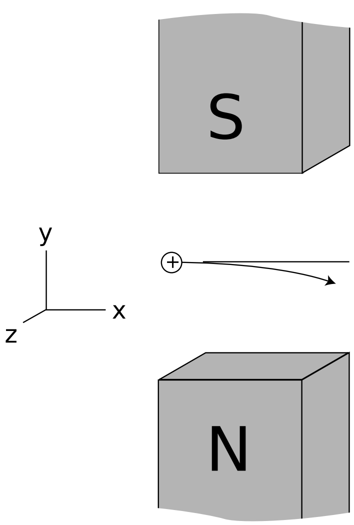
When it comes to electricity and magnetism, however, we have a problem, which was first clearly articulated by Einstein: if we state that magnetism is an interaction between moving charges, we have apparently created a law of physics that violates the principle that motion is relative, since different observers in different frames would disagree about how fast the charges were moving, or even whether they were moving at all. The incorrect solution that Einstein was taught (and disbelieved) as a student around the year 1900 was that the relative nature of motion applied only to mechanics, not to electricity and magnetism. The full story of how Einstein restored the principle of relative motion to its rightful place in physics involves his theory of special relativity, which we will not take up until book 6 of this series. However, a few simple and qualitative thought experiments will suffice to show how, based on the principle that motion is relative, there must be some new and previously unsuspected relationships between electricity and magnetism. These relationships form the basis for many practical, everyday devices, such as generators and transformers, and they also lead to an explanation of light itself as an electromagnetic phenomenon.

k / A line of positive charges.

Let's imagine an electrical example of relative motion in the same spirit as the story of the child in the back of the car. Suppose we have a line of positive charges, k. Observer A is in a frame of reference which is at rest with respect to these charges, and observes that they create an electric field pattern that points outward, away from the charges, in all directions, like a bottle brush. Suppose, however, that observer B is moving to the right with respect to the charges. As far as B is concerned, she's the one at rest, while the charges (and observer A) move to the left. In agreement with A, she observes an electric field, but since to her the charges are in motion, she must also observe a magnetic field in the same region of space, exactly like the magnetic field made by a current in a long, straight wire.

Who's right? They're both right. In A's frame of reference, there is only an **E**, while in B's frame there is both an **E** and a **B**. The principle of relative motion forces us to conclude that depending on our frame of reference we will observe a different combination of fields. Although we will not prove it (the proof requires special relativity, which we get to in book 6), it is true that either frame of reference provides a perfectly self-consistent description of things. For instance, if an electron passes through this region of space, both A and B will see it swerve, speed up, and slow down. A will successfully explain this as the result of an electric field, while B will ascribe the electron's behavior to a combination of electric and magnetic forces.

Thus, if we believe in the principle of relative motion, then we must accept that electric and magnetic fields are closely related phenomena, two sides of the same coin.




l / Observer A sees a positively charged particle moves through a region of upward magnetic field, which we assume to be uniform, between the poles of two magnets. The resulting force along the *z* axis causes the particle's path to curve toward us.

Now consider figure l. Observer A is at rest with respect to the bar magnets, and sees the particle swerving off in the $z$ direction, as it should according to the rule given in section 6.2 (sighting along the force vector, i.e., from behind the page, the **B** vector is clockwise from the $v$ vector). Suppose observer B, on the other hand, is moving to the right along the $x$ axis, initially at the same speed as the particle. B sees the bar magnets moving to the left and the particle initially at rest but then accelerating along the $z$ axis in a straight line. It is not possible for a magnetic field to start a particle moving if it is initially at rest, since magnetism is an interaction of moving charges with moving charges. B is thus led to the inescapable conclusion that there is an electric field in this region of space, which points along the $z$ axis. In other words, what A perceives as a pure **B** field, B sees as a mixture of **E** and **B**.

In general, observers who are not at rest with respect to one another will perceive different mixtures of electric and magnetic fields.

### The principle of induction

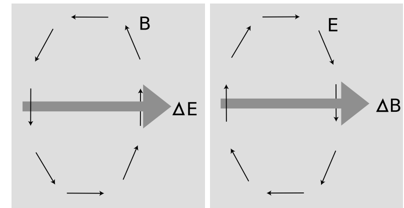
So far everything we've been doing might not seem terribly useful, since it seems that nothing surprising will happen as long as we stick to a single frame of reference, and don't worry about what people in other frames think. That isn't the whole story, however, as was discovered experimentally by Faraday in 1831 and explored mathematically by Maxwell later in the same century. Let's state Faraday's idea first, and then see how something like it must follow inevitably from the principle that motion is relative:

**the principle of induction**

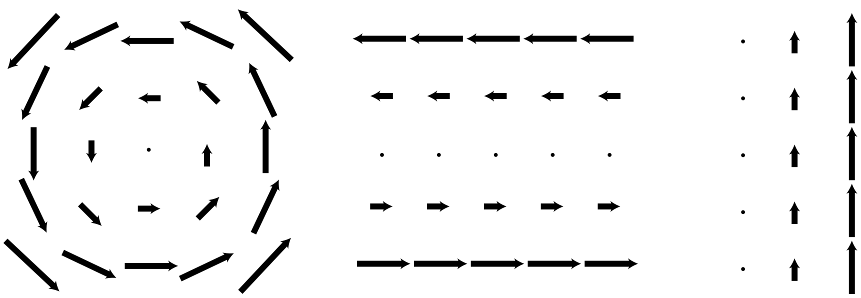
Any electric field that changes over time will produce a magnetic field in the space around it.

Any magnetic field that changes over time will produce an electric field in the space around it.

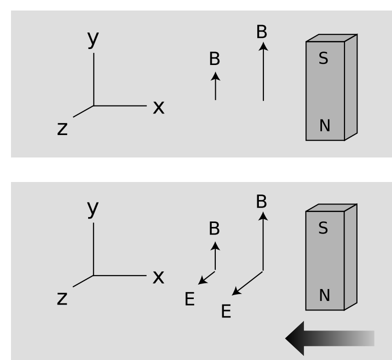
The induced field tends to have a whirlpool pattern, as shown in figure m, but the whirlpool image is not to be taken too literally; the principle of induction really just requires a field pattern such that, if one inserted a paddlewheel in it, the paddlewheel would spin. All of the field patterns shown in figure n are ones that could be created by induction; all have a counterclockwise "curl" to them.




m / The geometry of induced fields. The induced field tends to form a whirlpool pattern around the change in the vector producing it. Note how they circulate in opposite directions.



n / Three fields with counterclockwise "curls."

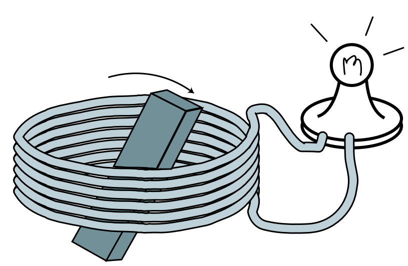



o / 1. Observer A is at rest with respect to the bar magnet, and observes magnetic fields that have different strengths at different distances from the magnet. 2. Observer B, hanging out in the region to the left of the magnet, sees the magnet moving toward her, and detects that the magnetic field in that region is getting stronger as time passes. As in 1, there is an electric field along the z axis because she's in motion with respect to the magnet. The $\Delta\mathbf{B}$ vector is upward, and the electric field has a curliness to it: a paddlewheel inserted in the electric field would spin clockwise as seen from above, since the clockwise torque made by the strong electric field on the right is greater than the counterclockwise torque made by the weaker electric field on the left.

Figure o shows an example of the fundamental reason why a changing **B** field must create an **E** field. The electric field would be inexplicable to observer B if she believed only in Coulomb's law, and thought that all electric fields are made by electric charges. If she knows about the principle of induction, however, the existence of this field is to be expected.

*the generator* *example 2*

A generator, p, consists of a permanent magnet that rotates within a coil of wire. The magnet is turned by a motor or crank, (not shown). As it spins, the nearby magnetic field changes. According to the principle of induction, this changing magnetic field results in an electric field, which has a whirlpool pattern. This electric field pattern creates a current that whips around the coils of wire, and we can tap this current to light the



p / A generator

lightbulb.

*self-check A*

When you're driving a car, the engine recharges the battery continuously using a device called an alternator, which is really just a generator like the one shown on the previous page, except that the coil rotates while the permanent magnet is fixed in place. Why can't you use the alternator to start the engine if your car's battery is dead? ▷ Answer, p. 196

*The transformer* *example 3*

In section 4.3 we discussed the advantages of transmitting power over electrical lines using high voltages and low currents. However, we don't want our wall sockets to operate at 10000 volts! For this reason, the electric company uses a device called a transformer, (g), to convert to lower voltages and higher currents inside your house. The coil on the input side creates a magnetic field. Transformers work with alternating current, so the magnetic field surrounding the input coil is always changing. This induces an electric field, which drives a current around the output coil.

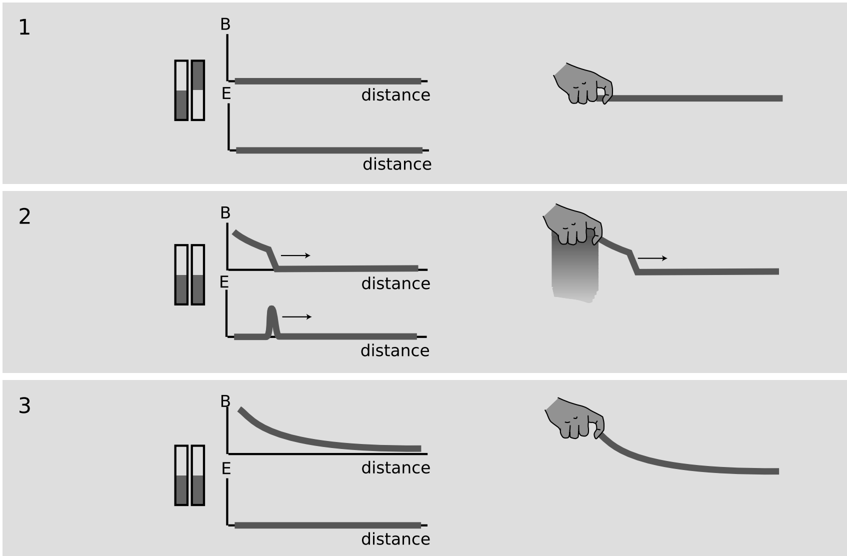
If both coils were the same, the arrangement would be symmetric, and the output would be the same as the input, but an output coil with a smaller number of coils gives the electric forces a smaller distance through which to push the electrons. Less mechanical work per unit charge means a lower voltage. Conservation of energy, however, guarantees that the amount of power on the output side must equal the amount put in originally, $I_{in}V_{in} = I_{out}V_{out}$, so this reduced voltage must be accompanied by an increased current.

*A mechanical analogy* *example 4*

Figure q shows an example of induction (left) with a mechanical analogy (right). The two bar magnets are initially pointing in opposite directions, 1, and their magnetic fields cancel out. If one magnet is flipped, 2, their fields reinforce, but the change in the magnetic field takes time to spread through space. Eventually, 3, the field becomes what you would expect from the theory of magnetostatics. In the mechanical analogy, the sudden motion of the hand produces a violent kink or wave pulse in the rope, the pulse travels along the rope, and it takes some time for the rope to settle down. An electric field is also induced in by the changing magnetic field, even though there is no net charge anywhere to to act as a source. (These simplified drawings are not meant to be accurate representations of the complete three-dimensional pattern of electric and magnetic fields.)

## Discussion Question

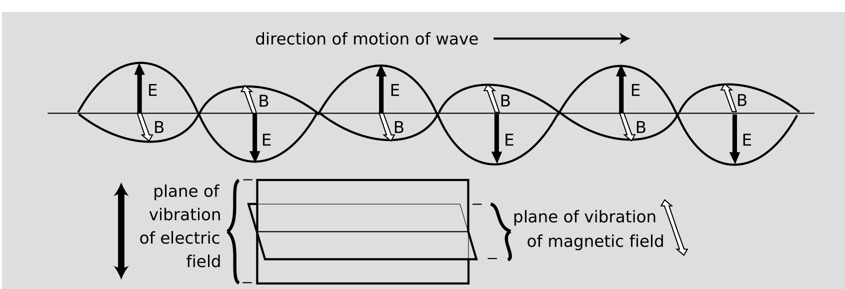
**A** In figures l and o, observer B is moving to the right. What would have happened if she had been moving to the left?




q / Example 4.

## 6.4 Electromagnetic Waves

The most important consequence of induction is the existence of electromagnetic waves. Whereas a gravitational wave would consist of nothing more than a rippling of gravitational fields, the principle of induction tells us that there can be no purely electrical or purely magnetic waves. Instead, we have waves in which there are both electric and magnetic fields, such as the sinusoidal one shown in the figure. Maxwell proved that such waves were a direct consequence of his equations, and derived their properties mathematically. The derivation would be beyond the mathematical level of this book, so we will just state the results.




r / An electromagnetic wave.

A sinusoidal electromagnetic wave has the geometry shown in figure r. The **E** and **B** fields are perpendicular to the direction of motion, and are also perpendicular to each other. If you look along the direction of motion of the wave, the **B** vector is always 90 degrees clockwise from the **E** vector. The magnitudes of the two fields are related by the equation $|\mathbf{E}| = c|\mathbf{B}|$.

How is an electromagnetic wave created? It could be emitted, for example, by an electron orbiting an atom or currents going back and forth in a transmitting antenna. In general any accelerating charge will create an electromagnetic wave, although only a current that varies sinusoidally with time will create a sinusoidal wave. Once created, the wave spreads out through space without any need for charges or currents along the way to keep it going. As the electric field oscillates back and forth, it induces the magnetic field, and the oscillating magnetic field in turn creates the electric field. The whole wave pattern propagates through empty space at a velocity $c = 3.0 \times 10^8$ m/s, which is related to the constants $k$ and $\mu_o$ by $c = \sqrt{4\pi k/\mu_o}$.

### Polarization

Two electromagnetic waves traveling in the same direction through space can differ by having their electric and magnetic fields in different directions, a property of the wave called its polarization.

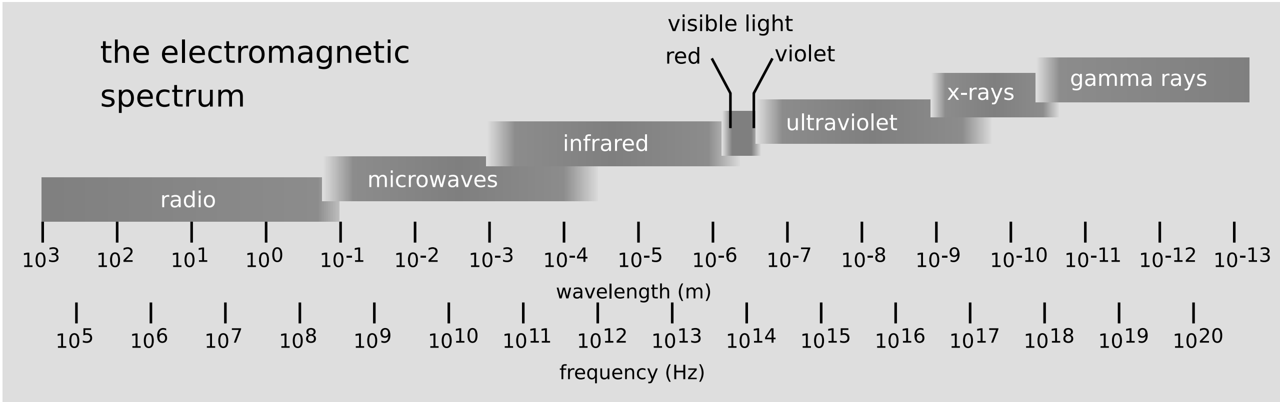
### Light is an electromagnetic wave

Once Maxwell had derived the existence of electromagnetic waves, he became certain that they were the same phenomenon as light. Both are transverse waves (i.e., the vibration is perpendicular to the direction the wave is moving), and the velocity is the same.

Heinrich Hertz (for whom the unit of frequency is named) verified Maxwell's ideas experimentally. Hertz was the first to succeed in producing, detecting, and studying electromagnetic waves in detail using antennas and electric circuits. To produce the waves, he had to make electric currents oscillate very rapidly in a circuit. In fact, there was really no hope of making the current reverse directions at the frequencies of $10^{15}$ Hz possessed by visible light. The fastest electrical oscillations he could produce were $10^9$ Hz, which would give a wavelength of about 30 cm. He succeeded in showing that, just like light, the waves he produced were polarizable, and could be reflected and refracted (i.e., bent, as by a lens), and he built devices such as parabolic mirrors that worked according to the same optical principles as those employing light. Hertz's results were convincing evidence that light and electromagnetic waves were one and the same.

### The electromagnetic spectrum

Today, electromagnetic waves with frequencies in the range employed by Hertz are known as radio waves. Any remaining doubts that the "Hertzian waves," as they were then called, were the same type of wave as light waves were soon dispelled by experiments in the whole range of frequencies in between, as well as the frequencies outside that range. In analogy to the spectrum of visible light, we speak of the entire electromagnetic spectrum, of which the visible spectrum is one segment.




The terminology for the various parts of the spectrum is worth memorizing, and is most easily learned by recognizing the logical relationships between the wavelengths and the properties of the waves with which you are already familiar. Radio waves have wavelengths that are comparable to the size of a radio antenna, i.e., meters to tens of meters. Microwaves were named that because they have much shorter wavelengths than radio waves; when food heats unevenly in a microwave oven, the small distances between neighboring hot and cold spots is half of one wavelength of the standing wave the oven creates. The infrared, visible, and ultraviolet obviously have much shorter wavelengths, because otherwise the wave nature of light would have been as obvious to humans as the wave nature of ocean waves. To remember that ultraviolet, x-rays, and gamma rays all lie on the short-wavelength side of visible, recall that all three of these can cause cancer. (As we'll discuss later in the course, there is a basic physical reason why the cancer-causing disruption of DNA can only be caused by very short-wavelength electromagnetic waves. Contrary to popular belief, microwaves cannot cause cancer, which is why we have microwave ovens and not x-ray ovens!)

s / Heinrich Hertz (1857-1894).

*Why the sky is blue* *example 5*

When sunlight enters the upper atmosphere, a particular air molecule finds itself being washed over by an electromagnetic wave of frequency $f$. The molecule's charged particles (nuclei and electrons) act like oscillators being driven by an oscillating force, and respond by vibrating at the same frequency $f$. Energy is sucked out of the incoming beam of sunlight and converted into the kinetic energy of the oscillating particles. However, these particles are accelerating, so they act like little radio antennas that put the energy back out as spherical waves of light that spread out in all directions. An object oscillating at a frequency $f$ has an acceleration proportional to $f^2$, and an accelerating charged particle creates an electromagnetic wave whose fields are proportional to its acceleration, so the field of the reradiated spherical wave is proportional to $f^2$. The energy of a field is proportional to the square of the field, so the energy of the reradiated is proportional to $f^4$. Since blue light has about twice the frequency of red light, this process is about $2^4 = 16$ times as strong for blue as for red, and that's why the sky is blue.

## 6.5 Calculating Energy in Fields

We have seen that the energy stored in a wave (actually the energy density) is typically proportional to the square of the wave's amplitude. Fields of force can make wave patterns, for which we might expect the same to be true. This turns out to be true not only for wave-like field patterns but for all fields:

$$\text{energy stored in the gravitational field per m}^3 = -\frac{1}{8\pi G}|\mathbf{g}|^2$$
$$\text{energy stored in the electric field per m}^3 = \frac{1}{8\pi k}|\mathbf{E}^2|$$
$$\text{energy stored in the magnetic field per m}^3 = \frac{1}{2\mu_o}|\mathbf{B}|^2$$

Although funny factors of $8\pi$ and the plus and minus signs may have initially caught your eye, they are not the main point. The important idea is that the energy density is proportional to the square of the field strength in all three cases. We first give a simple numerical example and work a little on the concepts, and then turn our attention to the factors out in front.

*Getting killed by a solenoid* *example 6*

Solenoids are very common electrical devices, but they can be a hazard to someone who is working on them. Imagine a solenoid that initially has a DC current passing through it. The current creates a magnetic field inside and around it, which contains energy. Now suppose that we break the circuit. Since there is no longer a complete circuit, current will quickly stop flowing, and the magnetic field will collapse very quickly. The field had energy stored in it, and even a small amount of energy can create a dangerous power surge if released over a short enough time interval. It is prudent not to fiddle with a solenoid that has current flowing through it, since breaking the circuit could be hazardous to your health.

As a typical numerical estimate, let's assume a 40 cm $\times$ 40 cm $\times$ 40 cm solenoid with an interior magnetic field of 1.0 T (quite a strong field). For the sake of this rough estimate, we ignore the exterior field, which is weak, and assume that the solenoid is cubical in shape. The energy stored in the field is

$$\begin{aligned}\text{(energy per unit volume)(volume)} &= \frac{1}{2\mu_o}|\mathbf{B}|^2 V \\ &= 3 \times 10^4 \text{ J}\end{aligned}$$

That's a lot of energy!

In chapter 5 when we discussed the original reason for introducing the concept of a field of force, a prime motivation was that otherwise there was no way to account for the energy transfers involved when forces were delayed by an intervening distance. We used to think of the universe's energy as consisting of

kinetic energy
+gravitational potential energy based on the distances between objects that interact gravitationally
+electric potential energy based on the distances between objects that interact electrically
+magnetic potential energy based on the distances between objects that interact magnetically ,

but in nonstatic situations we must use a different method:

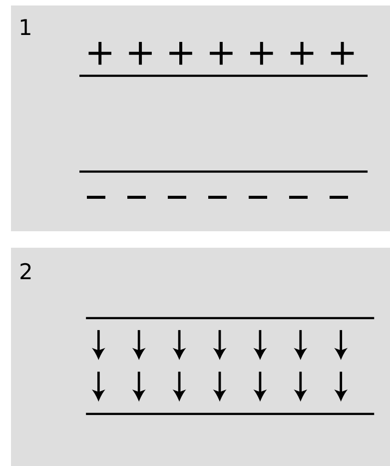
kinetic energy
+gravitational potential energy stored in gravitational fields
+electric potential energy stored in electric fields
+magnetic potential stored in magnetic fields

Surprisingly, the new method still gives the same answers for the static cases.

*Energy stored in a capacitor* *example 7*

A pair of parallel metal plates, seen from the side in figure t, can be used to store electrical energy by putting positive charge on one side and negative charge on the other. Such a device is called a capacitor. (We have encountered such an arrangement previously, but there its purpose was to deflect a beam of electrons, not to store energy.)

In the old method of describing potential energy, 1, we think in terms of the mechanical work that had to be done to separate the positive and negative charges onto the two plates, working against their electrical attraction. The new description, 2, attributes the storage of energy to the newly created electric field occupying the volume between the plates. Since this is a static case, both methods give the same, correct answer.



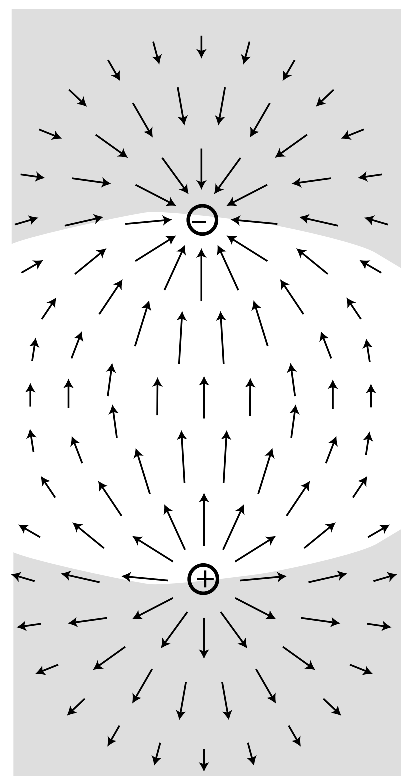

t / Example 7.

*Potential energy of a pair of opposite charges* *example 8*

Imagine taking two opposite charges, u, that were initially far apart and allowing them to come together under the influence of their electrical attraction.

According to the old method, potential energy is lost because the electric force did positive work as it brought the charges together. (This makes sense because as they come together and accelerate it is their potential energy that is being lost and converted to kinetic energy.)

By the new method, we must ask how the energy stored in the electric field has changed. In the region indicated approximately by the shading in the figure, the superposing fields of the two charges undergo partial cancellation because they are in opposing directions. The energy in the



u / Example 8.

shaded region is reduced by this effect. In the unshaded region, the fields reinforce, and the energy is increased.

It would be quite a project to do an actual numerical calculation of the energy gained and lost in the two regions (this is a case where the old method of finding energy gives greater ease of computation), but it is fairly easy to convince oneself that the energy is less when the charges are closer. This is because bringing the charges together shrinks the high-energy unshaded region and enlarges the low-energy shaded region.

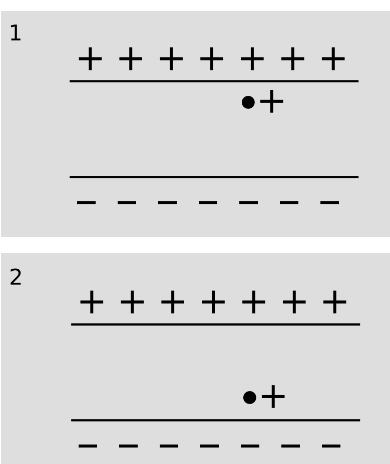
*Energy in an electromagnetic wave* — *example 9*

The old method would give zero energy for a region of space containing an electromagnetic wave but no charges. That would be wrong! We can only use the old method in static cases.

Now let's give at least some justification for the other features of the three expressions for energy density, $-\frac{1}{8\pi G}|\mathbf{g}|^2$, $\frac{1}{8\pi k}|\mathbf{E}^2|$, and $\frac{1}{2\mu_o}|\mathbf{B}|^2$, besides the proportionality to the square of the field strength.

First, why the different plus and minus signs? The basic idea is that the signs have to be opposite in the gravitational and electric cases because there is an attraction between two positive masses (which are the only kind that exist), but two positive charges would repel. Since we've already seen examples where the positive sign in the electric energy makes sense, the gravitational energy equation must be the one with the minus sign.

It may also seem strange that the constants $G, k$, and $\mu_o$ are in the denominator. They tell us how strong the three different forces are, so shouldn't they be on top? No. Consider, for instance, an alternative universe in which gravity is twice as strong as in ours. The numerical value of $G$ is doubled. Because $G$ is doubled, all the gravitational field strengths are doubled as well, which quadruples the quantity . In the expression $-\frac{1}{8\pi G}|\mathbf{g}|^2$, we have quadrupled something on top and doubled something on the bottom, which makes the energy twice as big. That makes perfect sense.



v / Discussion question A.

## Discussion Questions

**A** The figure shows a positive charge in the gap between two capacitor plates. First make a large drawing of the field pattern that would be formed by the capacitor itself, without the extra charge in the middle. Next, show how the field pattern changes when you add the particle at these two positions. Compare the energy of the electric fields in the two cases. Does this agree with what you would have expected based on your knowledge of electrical forces?

**B** Criticize the following statement: "A solenoid makes a charge in the space surrounding it, which dissipates when you release the energy."

**C** In the example on the previous page, I argued that the fields surrounding a positive and negative charge contain less energy when the charges are closer together. Perhaps a simpler approach is to consider the two extreme possibilities: the case where the charges are infinitely far apart, and the one in which they are at zero distance from each other, i.e., right on top of each other. Carry out this reasoning for the case of (1) a positive charge and a negative charge of equal magnitude, (2) two positive charges of equal magnitude, (3) the gravitational energy of two equal masses.

## 6.6 ⋆ Symmetry and Handedness

The physicist Richard Feynman helped to get me hooked on physics with an educational film containing the following puzzle. Imagine that you establish radio contact with an alien on another planet. Neither of you even knows where the other one's planet is, and you aren't able to establish any landmarks that you both recognize. You manage to learn quite a bit of each other's languages, but you're stumped when you try to establish the definitions of left and right (or, equivalently, clockwise and counterclockwise). Is there any way to do it?

If there was any way to do it without reference to external landmarks, then it would imply that the laws of physics themselves were asymmetric, which would be strange. Why should they distinguish left from right? The gravitational field pattern surrounding a star or planet looks the same in a mirror, and the same goes for electric fields. However, the field patterns shown in section 6.2 seem to violate this principle, but do they really? Could you use these patterns to explain left and right to the alien? In fact, the answer is no. If you look back at the definition of the magnetic field in section 6.1, it also contains a reference to handedness: the counterclockwise direction of the loop's current as viewed along the magnetic field. The aliens might have reversed their definition of the magnetic field, in which case their drawings of field patterns would look like mirror images of ours.

Until the middle of the twentieth century, physicists assumed that any reasonable set of physical laws would have to have this kind of symmetry between left and right. An asymmetry would be grotesque. Whatever their aesthetic feelings, they had to change their opinions about reality when experiments showed that the weak nuclear force (section 6.5) violates right-left symmetry! It is still a mystery why right-left symmetry is observed so scrupulously in general, but is violated by one particular type of physical process.

## Summary

### Selected Vocabulary

magnetic field . . . a field of force, defined in terms of the torque exerted on a test dipole

magnetic dipole . an object, such as a current loop, an atom, or a bar magnet, that experiences torques due to magnetic forces; the strength of magnetic dipoles is measured by comparison with a standard dipole consisting of a square loop of wire of a given size and carrying a given amount of current

induction . . . . . the production of an electric field by a changing magnetic field, or vice-versa

### Notation

**B** . . . . . . . . . the magnetic field

$D_m$ . . . . . . . . magnetic dipole moment

### Summary

Magnetism is an interaction of moving charges with other moving charges. The magnetic field is defined in terms of the torque on a magnetic test dipole. It has no sources or sinks; magnetic field patterns never converge on or diverge from a point.

The magnetic and electric fields are intimately related. The principle of induction states that any changing electric field produces a magnetic field in the surrounding space, and vice-versa. These induced fields tend to form whirlpool patterns.

The most important consequence of the principle of induction is that there are no purely magnetic or purely electric waves. Disturbances in the electrical and magnetic fields propagate outward as combined magnetic and electric waves, with a well-defined relationship between their magnitudes and directions. These electromagnetic waves are what light is made of, but other forms of electromagnetic waves exist besides visible light, including radio waves, x-rays, and gamma rays.

Fields of force contain energy. The density of energy is proportional to the square of the magnitude of the field. In the case of static fields, we can calculate potential energy either using the previous definition in terms of mechanical work or by calculating the energy stored in the fields. If the fields are not static, the old method gives incorrect results and the new one must be used.

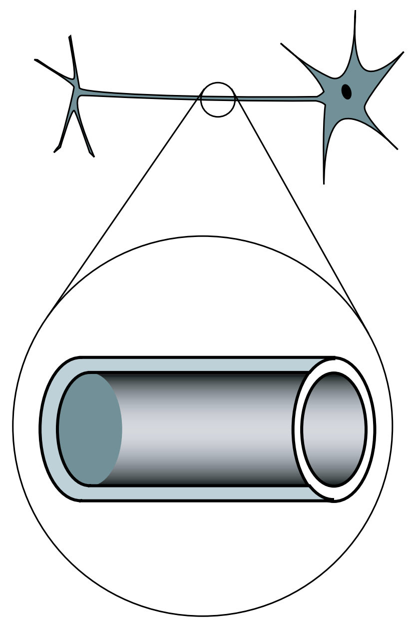
## Problems

### Key

- √ A computerized answer check is available online.
- ∫ A problem that requires calculus.
- ⋆ A difficult problem.

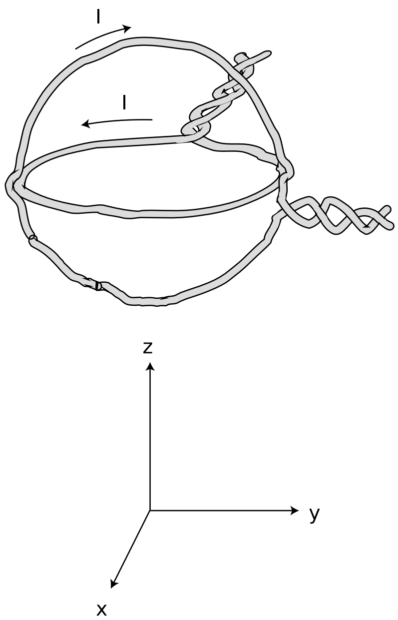
**1** In an electrical storm, the cloud and the ground act like a parallel-plate capacitor, which typically charges up due to frictional electricity in collisions of ice particles in the cold upper atmosphere. Lightning occurs when the magnitude of the electric field builds up to a critical value, $E_c$, at which air is ionized.
(a) Treat the cloud as a flat square with sides of length $L$. If it is at a height $h$ above the ground, find the amount of energy released in the lightning strike. √
(b) Based on your answer from part a, which is more dangerous, a lightning strike from a high-altitude cloud or a low-altitude one?
(c) Make an order-of-magnitude estimate of the energy released by a typical lightning bolt, assuming reasonable values for its size and altitude. $E_c$ is about $10^6$ V/m.

See problem 21 for a note on how recent research affects this estimate.


Problem 2.

**2** The neuron in the figure has been drawn fairly short, but some neurons in your spinal cord have tails (axons) up to a meter long. The inner and outer surfaces of the membrane act as the "plates" of a capacitor. (The fact that it has been rolled up into a cylinder has very little effect.) In order to function, the neuron must create a voltage difference $V$ between the inner and outer surfaces of the membrane. Let the membrane's thickness, radius, and length be $t$, $r$, and $L$. (a) Calculate the energy that must be stored in the electric field for the neuron to do its job. (In real life, the membrane is made out of a substance called a dielectric, whose electrical properties increase the amount of energy that must be stored. For the sake of this analysis, ignore this fact.) [Hint: The volume of the membrane is essentially the same as if it was unrolled and flattened out.] √
(b) An organism's evolutionary fitness should be better if it needs less energy to operate its nervous system. Based on your answer to part a, what would you expect evolution to do to the dimensions $t$ and $r$? What other constraints would keep these evolutionary trends from going too far?

**3** Consider two solenoids, one of which is smaller so that it can be put inside the other. Assume they are long enough so that each one only contributes significantly to the field inside itself, and the interior fields are nearly uniform. Consider the configuration where the small one is inside the big one with their currents circulating in the same direction, and a second configuration in which the currents circulate in opposite directions. Compare the energies of these configurations with the energy when the solenoids are far apart. Based on this reasoning, which configuration is stable, and in which configuration will the little solenoid tend to get twisted around or spit out? [Hint: A stable system has low energy; energy would have to be added to change its configuration.]




Problem 4.

**4** The figure shows a nested pair of circular wire loops used to create magnetic fields. (The twisting of the leads is a practical trick for reducing the magnetic fields they contribute, so the fields are very nearly what we would expect for an ideal circular current loop.) The coordinate system below is to make it easier to discuss directions in space. One loop is in the $y-z$ plane, the other in the $x-y$ plane. Each of the loops has a radius of 1.0 cm, and carries 1.0 A in the direction indicated by the arrow.

(a) Using the equation in optional section 6.2, calculate the magnetic field that would be produced by *one* such loop, at its center.

(b) Describe the direction of the magnetic field that would be produced, at its center, by the loop in the $x-y$ plane alone.

(c) Do the same for the other loop.

(d) Calculate the magnitude of the magnetic field produced by the two loops in combination, at their common center. Describe its direction. √ √

**5** (a) Show that the quantity $\sqrt{4\pi k/\mu_o}$ has units of velocity.

(b) Calculate it numerically and show that it equals the speed of light.

(c) Prove that in an electromagnetic wave, half the energy is in the electric field and half in the magnetic field.

**6** One model of the hydrogen atom has the electron circling around the proton at a speed of $2.2\times10^6$ m/s, in an orbit with a radius of 0.05 nm. (Although the electron and proton really orbit around their common center of mass, the center of mass is very close to the proton, since it is 2000 times more massive. For this problem, assume the proton is stationary.) In that previous homework problem, you calculated the electric current created.
(a) Now estimate the magnetic field created at the center of the atom by the electron. We are treating the circling electron as a current loop, even though it's only a single particle. √
(b) Does the proton experience a nonzero force from the electron's magnetic field? Explain.
(c) Does the electron experience a magnetic field from the proton? Explain.
(d) Does the electron experience a magnetic field created by its own current? Explain.
(e) Is there an electric force acting between the proton and electron? If so, calculate it. √
(f) Is there a gravitational force acting between the proton and elec-

tron? If so, calculate it.

(g) An inward force is required to keep the electron in its orbit – otherwise it would obey Newton's first law and go straight, leaving the atom. Based on your answers to the previous parts, which force or forces (electric, magnetic and gravitational) contributes significantly to this inward force?

**7** [You need to have read optional section 6.2 to do this problem.] Suppose a charged particle is moving through a region of space in which there is an electric field perpendicular to its velocity vector, and also a magnetic field perpendicular to both the particle's velocity vector and the electric field. Show that there will be one particular velocity at which the particle can be moving that results in a total force of zero on it. Relate this velocity to the magnitudes of the electric and magnetic fields. (Such an arrangement, called a velocity filter, is one way of determining the speed of an unknown particle.)

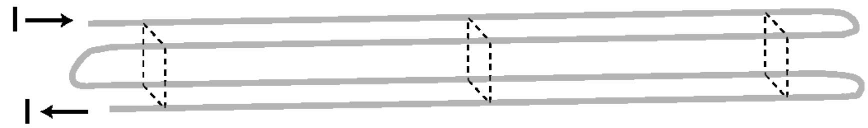
**8** If you put four times more current through a solenoid, how many times more energy is stored in its magnetic field?

**9** Suppose we are given a permanent magnet with a complicated, asymmetric shape. Describe how a series of measurements with a magnetic compass could be used to determine the strength and direction of its magnetic field at some point of interest. Assume that you are only able to see the direction to which the compass needle settles; you cannot measure the torque acting on it. $\star$

**10** Consider two solenoids, one of which is smaller so that it can be put inside the other. Assume they are long enough to act like ideal solenoids, so that each one only contributes significantly to the field inside itself, and the interior fields are nearly uniform. Consider the configuration where the small one is partly inside and partly hanging out of the big one, with their currents circulating in the same direction. Their axes are constrained to coincide.

(a) Find the magnetic potential energy as a function of the length $x$ of the part of the small solenoid that is inside the big one. (Your equation will include other relevant variables describing the two solenoids.)

(b) Based on your answer to part (a), find the force acting between the solenoids.




Problem 11.

**11** Four long wires are arranged, as shown, so that their cross-section forms a square, with connections at the ends so that current flows through all four before exiting.

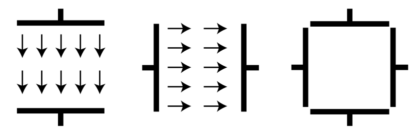
Note that the current is to the right in the two back wires, but to the left in the front wires. If the dimensions of the cross-sectional square (height and front-to-back) are $b$, find the magnetic field (magnitude and direction) along the long central axis.

**12** To do this problem, you need to understand how to do volume integrals in cylindrical and spherical coordinates. (a) Show that if you try to integrate the energy stored in the field of a long, straight wire, the resulting energy per unit length diverges both at $r \to 0$ and $r \to \infty$. Taken at face value, this would imply that a certain real-life process, the initiation of a current in a wire, would be impossible, because it would require changing from a state of zero magnetic energy to a state of infinite magnetic energy. (b) Explain why the infinities at $r \to 0$ and $r \to \infty$ don't really happen in a realistic situation. (c) Show that the electric energy of a point charge diverges at $r \to 0$, but not at $r \to \infty$.

A remark regarding part (c): Nature does seem to supply us with particles that are charged and pointlike, e.g., the electron, but one could argue that the infinite energy is not really a problem, because an electron traveling around and doing things neither gains nor loses infinite energy; only an infinite *change* in potential energy would be physically troublesome. However, there are real-life processes that create and destroy pointlike charged particles, e.g., the annihilation of an electron and antielectron with the emission of two gamma rays. Physicists have, in fact, been struggling with infinities like this since about 1950, and the issue is far from resolved. Some theorists propose that apparently pointlike particles are actually not pointlike: close up, an electron might be like a little circular loop of string. $\int$ $\star$

**13** The purpose of this problem is to find the force experienced by a straight, current-carrying wire running perpendicular to a uniform magnetic field. (a) Let $A$ be the cross-sectional area of the wire, $n$ the number of free charged particles per unit volume, $q$ the charge per particle, and $v$ the average velocity of the particles. Show that the current is $I = Avnq$. (b) Show that the magnetic force per unit length is $AvnqB$. (c) Combining these results, show that the force on the wire per unit length is equal to $IB$. ▷ Solution, p. 199

**14** Suppose two long, parallel wires are carrying current $I_1$ and $I_2$. The currents may be either in the same direction or in opposite directions. (a) Using the information from section 6.2, determine under what conditions the force is attractive, and under what conditions it is repulsive. Note that, because of the difficulties explored in problem 12, it's possible to get yourself tied up in knots if you use the energy approach of section 6.5. (b) Starting from the result of problem 13, calculate the force per unit length. ▷ Solution, p. 199



Problem 15.

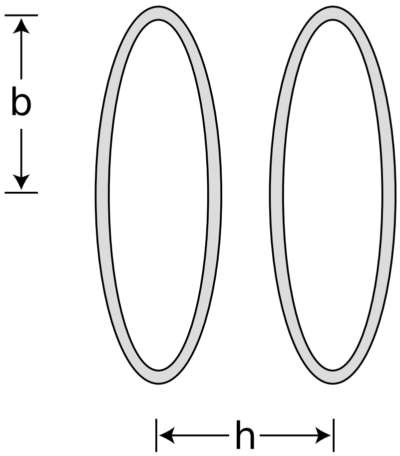
**15** The figure shows cross-sectional views of two cubical capacitors, and a cross-sectional view of the same two capacitors put together so that their interiors coincide. A capacitor with the plates close together has a nearly uniform electric field between the plates, and almost zero field outside; these capacitors don't have their plates very close together compared to the dimensions of the plates, but for the purposes of this problem, assume that they still have approximately the kind of idealized field pattern shown in the figure. Each capacitor has an interior volume of 1.00 $m^3$, and is charged up to the point where its internal field is 1.00 V/m. (a) Calculate the energy stored in the electric field of each capacitor when they are separate. (b) Calculate the magnitude of the interior field when the two capacitors are put together in the manner shown. Ignore effects arising from the redistribution of each capacitor's charge under the influence of the other capacitor. (c) Calculate the energy of the put-together configuration. Does assembling them like this release energy, consume energy, or neither?

**16** Section 6.2 states the following rule:

For a positively charged particle, the direction of the $F$ vector is the one such that if you sight along it, the $\mathbf{B}$ vector is clockwise from the $v$ vector.

Make a three-dimensional model of the three vectors using pencils or rolled-up pieces of paper to represent the vectors assembled with their tails together. Now write down every possible way in which the rule could be rewritten by scrambling up the three symbols $F$, $\mathbf{B}$, and $v$. Referring to your model, which are correct and which are incorrect?

**17** Prove that any two planar current loops with the same value of $IA$ will experience the same torque in a magnetic field, regardless of their shapes. In other words, the dipole moment of a current loop can be defined as $IA$, regardless of whether its shape is a square. ⋆




Problem 18.

**18** A Helmholtz coil is defined as a pair of identical circular coils separated by a distance, $h$, equal to their radius, $b$. (Each coil may have more than one turn of wire.) Current circulates in the same direction in each coil, so the fields tend to reinforce each other in the interior region. This configuration has the advantage of being fairly open, so that other apparatus can be easily placed inside and subjected to the field while remaining visible from the outside. The choice of $h = b$ results in the most uniform possible field near the center. (a) Find the percentage drop in the field at the center of one coil, compared to the full strength at the center of the whole apparatus. (b) What value of $h$ (not equal to $b$) would make this percentage difference equal to zero?

**19** (a) In the photo of the vacuum tube apparatus in section 6.2, infer the direction of the magnetic field from the motion of the electron beam. (b) Based on your answer to a, find the direction of the currents in the coils. (c) What direction are the electrons in the coils going? (d) Are the currents in the coils repelling or attracting the currents consisting of the beam inside the tube? Compare with part a of problem 14.

**20** In the photo of the vacuum tube apparatus in section 6.2, an approximately uniform magnetic field caused circular motion. Is there any other possibility besides a circle? What can happen in general? $\star$

**21** In problem 1, you estimated the energy released in a bolt of lightning, based on the energy stored in the electric field immediately before the lightning occurs. The assumption was that the field would build up to a certain value, which is what is necessary to ionize air. However, real-life measurements always seemed to show electric fields strengths roughtly 10 times smaller than those required in that model. For a long time, it wasn't clear whether the field measurements were wrong, or the model was wrong. Research carried out in 2003 seems to show that the model was wrong. It is now believed that the final triggering of the bolt of lightning comes from cosmic rays that enter the atmosphere and ionize some of the air. If the field is 10 times smaller than the value assumed in problem 1, what effect does this have on the final result of problem 1?

**22** In section 6.2 I gave an equation for the magnetic field in the interior of a solenoid, but that equation doesn't give the right answer near the mouths or on the outside. Although in general the computation of the field in these other regions is complicated, it is possible to find a precise, simple result for the field at the center of one of the mouths, using only symmetry and vector addition. What is it? ▷ Solution, p. 200 $\star$

# Chapter A
# Capacitance and Inductance

*This chapter is optional.*

The long road leading from the light bulb to the computer started with one very important step: the introduction of feedback into electronic circuits. Although the principle of feedback has been understood and and applied to mechanical systems for centuries, and to electrical ones since the early twentieth century, for most of us the word evokes an image of Jimi Hendrix (or some more recent guitar hero) intentionally creating earsplitting screeches, or of the school principal doing the same inadvertently in the auditorium. In the guitar example, the musician stands in front of the amp and turns it up so high that the sound waves coming from the speaker come back to the guitar string and make it shake harder. This is an example of *positive* feedback: the harder the string vibrates, the stronger the sound waves, and the stronger the sound waves, the harder the string vibrates. The only limit is the power-handling ability of the amplifier.

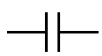
Negative feedback is equally important. Your thermostat, for example, provides negative feedback by kicking the heater off when the house gets warm enough, and by firing it up again when it gets too cold. This causes the house's temperature to oscillate back and forth within a certain range. Just as out-of-control exponential freak-outs are a characteristic behavior of positive-feedback systems, oscillation is typical in cases of negative feedback. You have already studied negative feedback extensively in *Vibrations and Waves* in the case of a mechanical system, although we didn't call it that.

## A.1 Capacitance and inductance

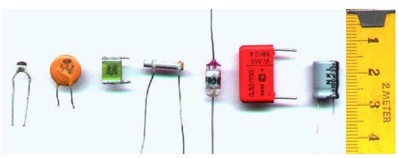
In a mechanical oscillation, energy is exchanged repetitively between potential and kinetic forms, and may also be siphoned off in the form of heat dissipated by friction. In an electrical circuit, resistors are the circuit elements that dissipate heat. What are the electrical analogs of storing and releasing the potential and kinetic energy of a vibrating object? When you think of energy storage in an electrical circuit, you are likely to imagine a battery, but even rechargeable batteries can only go through 10 or 100 cycles before they wear out.

In addition, batteries are not able to exchange energy on a short enough time scale for most applications. The circuit in a musical synthesizer may be called upon to oscillate thousands of times a second, and your microwave oven operates at gigahertz frequencies. Instead of batteries, we generally use capacitors and inductors to store energy in oscillating circuits. Capacitors, which you've already encountered, store energy in electric fields. An inductor does the same with magnetic fields.

### Capacitors



a / The symbol for a capacitor.



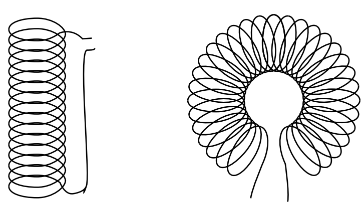
b / Some capacitors.

A capacitor's energy exists in its surrounding electric fields. It is proportional to the square of the field strength, which is proportional to the charges on the plates. If we assume the plates carry charges that are the same in magnitude, $+q$ and $-q$, then the energy stored in the capacitor must be proportional to $q^2$. For historical reasons, we write the constant of proportionality as $1/2C$,

$$E_C = \frac{1}{2C}q^2 \qquad .$$

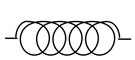
The constant $C$ is a geometrical property of the capacitor, called its capacitance.

Based on this definition, the units of capacitance must be coulombs squared per joule, and this combination is more conveniently abbreviated as the farad, $1\ \text{F} = 1\ \text{C}^2/\text{J}$. "Condenser" is a less formal term for a capacitor. Note that the labels printed on capacitors often use MF to mean $\mu$F, even though MF should really be the symbol for megafarads, not microfarads. Confusion doesn't result from this nonstandard notation, since picofarad and microfarad values are the most common, and it wasn't until the 1990's that even millifarad and farad values became available in practical physical sizes. Figure a show the symbol used in schematics to represent a capacitor.

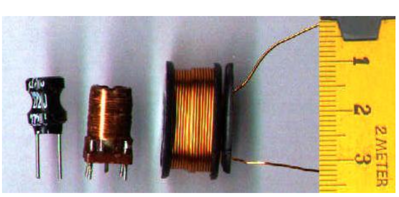


c / Two common geometries for inductors. The cylindrical shape on the left is called a solenoid.

### Inductors



d / The symbol for an inductor.



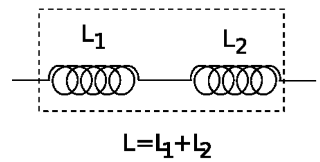
e / Some inductors.

Any current will create a magnetic field, so in fact every current-carrying wire in a circuit acts as an inductor! However, this type of "stray" inductance is typically negligible, just as we can usually ignore the stray resistance of our wires and only take into account the actual resistors. To store any appreciable amount of magnetic energy, one usually uses a coil of wire designed specifically to be an inductor. All the loops' contribution to the magnetic field add together to make a stronger field. Unlike capacitors and resistors, practical inductors are easy to make by hand. One can for instance spool some wire around a short wooden dowel, put the spool inside a plastic aspirin bottle with the leads hanging out, and fill the bottle with epoxy to make the whole thing rugged. An inductor like this, in the form cylindrical coil of wire, is called a solenoid, c, and a stylized solenoid, d, is the symbol used to represent an inductor in a circuit regardless of its actual geometry.

How much energy does an inductor store? The energy density is proportional to the square of the magnetic field strength, which is in turn proportional to the current flowing through the coiled wire, so the energy stored in the inductor must be proportional to $I^2$. We write $L/2$ for the constant of proportionality, giving

$$E_L = \frac{L}{2} I^2 \qquad .$$

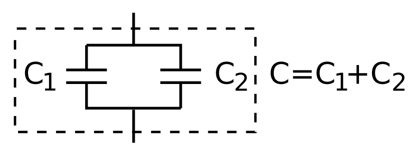
As in the definition of capacitance, we have a factor of 1/2, which is purely a matter of definition. The quantity $L$ is called the *inductance* of the inductor, and we see that its units must be joules per ampere squared. This clumsy combination of units is more commonly abbreviated as the henry, 1 henry = 1 $\text{J/A}^2$. Rather than memorizing this definition, it makes more sense to derive it when needed from the definition of inductance. Many people know inductors simply as "coils," or "chokes," and will not understand you if you refer to an "inductor," but they will still refer to $L$ as the "inductance," not the "coilance" or "chokeance!"




f / Inductances in series add.

*Identical inductances in series* *example 1*

If two inductors are placed in series, any current that passes through the combined double inductor must pass through both its parts. Thus by the definition of inductance, the inductance is doubled as well. In general, inductances in series add, just like resistances. The same kind of reasoning also shows that the inductance of a solenoid is approximately proportional to its length, assuming the number of turns per unit length is kept constant.



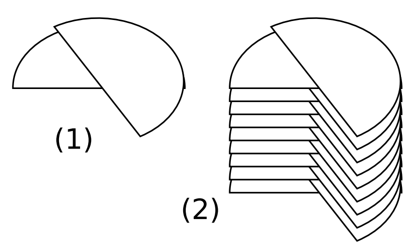

g / Capacitances in parallel add.

*Identical capacitances in parallel* *example 2*

When two identical capacitances are placed in parallel, any charge deposited at the terminals of the combined double capacitor will divide itself evenly between the two parts. The electric fields surrounding each capacitor will be half the intensity, and therefore store one quarter the energy. Two capacitors, each storing one quarter the energy, give half the total energy storage. Since capacitance is inversely related to energy storage, this implies that identical capacitances in parallel give double the capacitance. In general, capacitances in parallel add. This is unlike the behavior of inductors and resistors, for which series configurations give addition.

This is consistent with the fact that the capacitance of a single parallel-plate capacitor proportional to the area of the plates. If we have two parallel-plate capacitors, and we combine them in parallel and bring them very close together side by side, we have produced a single capacitor with plates of double the area, and it has approximately double the capacitance.

Inductances in parallel and capacitances in series are explored in homework problems 4 and 6.




h / A variable capacitor.

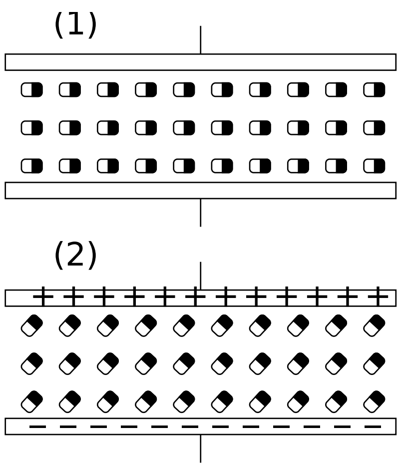
*A variable capacitor* *example 3*

Figure h/1 shows the construction of a variable capacitor out of two parallel semicircles of metal. One plate is fixed, while the other can be rotated about their common axis with a knob. The opposite charges on the two plates are attracted to one another, and therefore tend to gather in the overlapping area. This overlapping area, then, is the only area that effectively contributes to the capacitance, and turning the knob changes the capacitance. The simple design can only provide very small capacitance values, so in practice one usually uses a bank of capacitors, wired in parallel, with all the moving parts on the same shaft.

### Discussion Questions

**A** Suppose that two parallel-plate capacitors are wired in parallel, and are placed very close together, side by side, so that their fields overlap. Will the resulting capacitance be too small, or too big? Could you twist the circuit into a different shape and make the effect be the other way around, or make the effect vanish? How about the case of two inductors in series?

**B** Most practical capacitors do not have an air gap or vacuum gap between the plates; instead, they have an insulating substance called a dielectric. We can think of the molecules in this substance as dipoles that are free to rotate (at least a little), but that are not free to move around, since it is a solid. The figure shows a highly stylized and unrealistic way of visualizing this. We imagine that all the dipoles are intially turned sideways, (1), and that as the capacitor is charged, they all respond by turning through a certain angle, (2). (In reality, the scene might be much more random, and the alignment effect much weaker.)



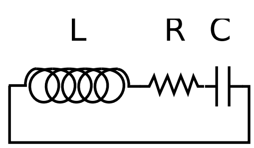
i / Discussion question B.

For simplicity, imagine inserting just one electric dipole into the vacuum gap. For a given amount of charge on the plates, how does this affect the amount of energy stored in the electric field? How does this affect the capacitance?

Now redo the analysis in terms of the mechanical work needed in order to charge up the plates.

## A.2 Oscillations

Figure j shows the simplest possible oscillating circuit. For any useful application it would actually need to include more components. For example, if it was a radio tuner, it would need to be connected to an antenna and an amplifier. Nevertheless, all the essential physics is there.

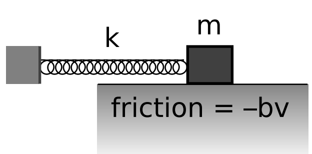


j / A series LRC circuit.

We can analyze it without any sweat or tears whatsoever, simply by constructing an analogy with a mechanical system. In a mechanical oscillator, k, we have two forms of stored energy,

$$E_{spring} = \frac{1}{2}kx^2 \qquad (1)$$

$$K = \frac{1}{2}mv^2 \qquad . \qquad (2)$$



k / A mechanical analogy for the LRC circuit.

In the case of a mechanical oscillator, we have usually assumed a friction force of the form that turns out to give the nicest mathematical results, $F = -bv$. In the circuit, the dissipation of energy into heat occurs via the resistor, with no mechanical force involved, so in order to make the analogy, we need to restate the role of the friction force in terms of energy. The power dissipated by friction equals the mechanical work it does in a time interval $\Delta t$, divided by $\Delta t$, $P = W/\Delta t = F\Delta x/\Delta t = Fv = -bv^2$, so

$$\text{rate of heat dissipation} = -bv^2 \quad . \qquad (3)$$

*self-check A*

Equation (1) has $x$ squared, and equations (2) and (3) have $v$ squared. Because they're squared, the results don't depend on whether these variables are positive or negative. Does this make physical sense? ▷ Answer, p. 196

In the circuit, the stored forms of energy are

$$E_C = \frac{1}{2C}q^2 \qquad (1')$$
$$E_L = \frac{1}{2}LI^2 \quad , \qquad (2')$$

and the rate of heat dissipation in the resistor is

$$\text{rate of heat dissipation} = -RI^2 \quad . \qquad (3')$$

Comparing the two sets of equations, we first form analogies between quantities that represent the state of the system at some moment in time:

$$x \leftrightarrow q$$
$$v \leftrightarrow I$$

*self-check B*

How is $v$ related mathematically to $x$? How is $I$ connected to $q$? Are the two relationships analogous? ▷ Answer, p. 196

Next we relate the ones that describe the system's permanent characteristics:

$$k \leftrightarrow 1/C$$
$$m \leftrightarrow L$$
$$b \leftrightarrow R$$

Since the mechanical system naturally oscillates with a period $T = 2\pi\sqrt{m/k}$ , we can immediately solve the electrical version by analogy, giving

$$T = 2\pi\sqrt{LC} \quad .$$

Rather than period, $T$, and frequency, $f$, it turns out to be more convenient if we work with the quantity $\omega = 2\pi f$, which can be interpreted as the number of radians per second. Then

$$\omega = \frac{1}{\sqrt{LC}} \quad .$$

Since the resistance $R$ is analogous to $b$ in the mechanical case, we find that the $Q$ (quality factor, not charge) of the resonance is inversely proportional to $R$, and the width of the resonance is directly proportional to $R$.

*Tuning a radio receiver* *example 4*

A radio receiver uses this kind of circuit to pick out the desired station. Since the receiver resonates at a particular frequency, stations whose frequencies are far off will not excite any response in the circuit. The value of $R$ has to be small enough so that only one station at a time is picked up, but big enough so that the tuner isn't too touchy. The resonant frequency can be tuned by adjusting either $L$ or $C$, but variable capacitors are easier to build than variable inductors.

*A numerical calculation* *example 5*

The phone company sends more than one conversation at a time over the same wire, which is accomplished by shifting each voice signal into different range of frequencies during transmission. The number of signals per wire can be maximized by making each range of frequencies (known as a bandwidth) as small as possible. It turns out that only a relatively narrow range of frequencies is necessary in order to make a human voice intelligible, so the phone company filters out all the extreme highs and lows. (This is why your phone voice sounds different from your normal voice.)

▷ If the filter consists of an LRC circuit with a broad resonance centered around 1.0 kHz, and the capacitor is 1 μF (microfarad), what inductance value must be used?

▷ Solving for $L$, we have

$$\begin{aligned} L &= \frac{1}{C\omega^2} \\ &= \frac{1}{(10^{-6}\ \text{F})(2\pi \times 10^3\ \text{s}^{-1})^2} \\ &= 2.5 \times 10^{-3}\ \text{F}^{-1}\text{s}^2 \end{aligned}$$

Checking that these really are the same units as henries is a little tedious, but it builds character:

$$\begin{aligned} \text{F}^{-1}\text{s}^2 &= (\text{C}^2/\text{J})^{-1}\text{s}^2 \\ &= \text{J}\cdot\text{C}^{-2}\text{s}^2 \\ &= \text{J}/\text{A}^2 \\ &= \text{H} \end{aligned}$$

The result is 25 mH (millihenries).

This is actually quite a large inductance value, and would require a big, heavy, expensive coil. In fact, there is a trick for making this kind of circuit small and cheap. There is a kind of silicon chip called an op-amp, which, among other things, can be used to simulate the behavior of an inductor. The main limitation of the op-amp is that it is restricted to low-power applications.

## A.3 Voltage and Current

What is physically happening in one of these oscillating circuits? Let's first look at the mechanical case, and then draw the analogy to the circuit. For simplicity, let's ignore the existence of damping, so there is no friction in the mechanical oscillator, and no resistance in the electrical one.

Suppose we take the mechanical oscillator and pull the mass away from equilibrium, then release it. Since friction tends to resist the spring's force, we might naively expect that having zero friction would allow the mass to leap instantaneously to the equilibrium position. This can't happen, however, because the mass would have to have infinite velocity in order to make such an instantaneous leap. Infinite velocity would require infinite kinetic energy, but the only kind of energy that is available for conversion to kinetic is the energy stored in the spring, and that is finite, not infinite. At each step on its way back to equilibrium, the mass's velocity is controlled exactly by the amount of the spring's energy that has so far been converted into kinetic energy. After the mass reaches equilibrium, it overshoots due to its own momentum. It performs identical oscillations on both sides of equilibrium, and it never loses amplitude because friction is not available to convert mechanical energy into heat.

Now with the electrical oscillator, the analog of position is charge. Pulling the mass away from equilibrium is like depositing charges $+q$ and $-q$ on the plates of the capacitor. Since resistance tends to resist the flow of charge, we might imagine that with no friction present, the charge would instantly flow through the inductor (which is, after all, just a piece of wire), and the capacitor would discharge instantly. However, such an instant discharge is impossible, because it would require infinite current for one instant. Infinite current would create infinite magnetic fields surrounding the inductor, and these fields would have infinite energy. Instead, the rate of flow of current is controlled at each instant by the relationship between the amount of energy stored in the magnetic field and the amount of current that must exist in order to have that strong a field. After the capacitor reaches $q = 0$, it overshoots. The circuit has its own kind of electrical "inertia," because if charge was to stop flowing, there would have to be zero current through the inductor. But the current in the inductor must be related to the amount of energy stored in its magnetic fields. When the capacitor is at $q = 0$, all the circuit's energy is in the inductor, so it must therefore have strong magnetic fields surrounding it and quite a bit of current going through it.

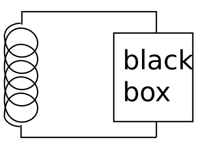
The only thing that might seem spooky here is that we used to speak as if the current in the inductor caused the magnetic field, but now it sounds as if the field causes the current. Actually this is symptomatic of the elusive nature of cause and effect in physics. It's equally valid to think of the cause and effect relationship in either way. This may seem unsatisfying, however, and for example does not really get at the question of what brings about a voltage difference across the resistor (in the case where the resistance is finite); there must be such a voltage difference, because without one, Ohm's law would predict zero current through the resistor.

Voltage, then, is what is really missing from our story so far.

Let's start by studying the voltage across a capacitor. Voltage is electrical potential energy per unit charge, so the voltage difference between the two plates of the capacitor is related to the amount by which its energy would increase if we increased the absolute values of the charges on the plates from $q$ to $q + \Delta q$:

$$\begin{aligned} V_C &= (E_{q+\Delta q} - E_q)/\Delta q \\ &= \frac{\Delta E_C}{\Delta q} \\ &= \frac{\Delta}{\Delta q}\left(\frac{1}{2C}q^2\right) \\ &= \frac{q}{C} \end{aligned}$$

Many books use this as the definition of capacitance. This equation, by the way, probably explains the historical reason why $C$ was defined so that the energy was *inversely* proportional to $C$ for a given value of $C$: the people who invented the definition were thinking of a capacitor as a device for storing charge rather than energy, and the amount of charge stored for a fixed voltage (the charge "capacity") is proportional to $C$.



l / The inductor releases energy and gives it to the black box.

In the case of an inductor, we know that if there is a steady, constant current flowing through it, then the magnetic field is constant, and so is the amount of energy stored; no energy is being exchanged between the inductor and any other circuit element. But what if the current is changing? The magnetic field is proportional to the current, so a change in one implies a change in the other. For concreteness, let's imagine that the magnetic field and the current are both decreasing. The energy stored in the magnetic field is therefore decreasing, and by conservation of energy, this energy can't just go away — some other circuit element must be taking energy from the inductor. The simplest example, shown in figure l, is a series circuit consisting of the inductor plus one other circuit element. It doesn't matter what this other circuit element is, so we just call it a black box, but if you like, we can think of it as a resistor, in which case the energy lost by the inductor is being turned into heat by the resistor. The junction rule tells us that both circuit elements have the same current through them, so $I$ could refer to either one, and likewise the loop rule tells us $V_{inductor} + V_{black\ box} = 0$, so the two voltage drops have the same absolute value, which we can refer to as $V$. Whatever the black box is, the rate at which it is taking energy from the inductor is given by $|P| = |IV|$, so

$$\begin{aligned}|IV| &= \left|\frac{\Delta E_L}{\Delta t}\right| \\ &= \left|\frac{\Delta}{\Delta t}\left(\frac{1}{2}LI^2\right)\right| \\ &= \left|LI\frac{\Delta I}{\Delta t}\right| \qquad ,\end{aligned}$$

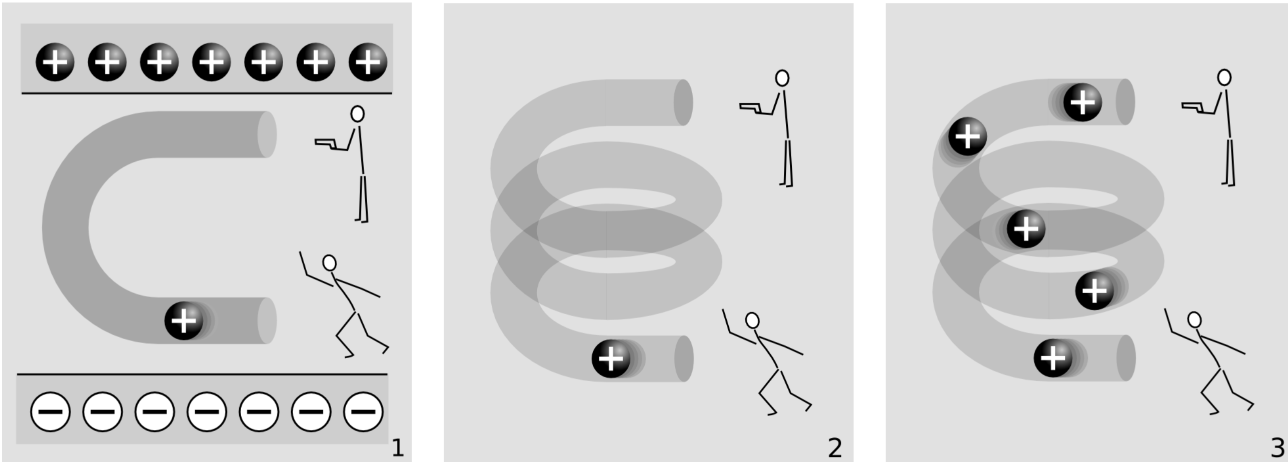
or

$$|V| = \left|L\frac{\Delta I}{\Delta t}\right| \qquad ,$$

which in many books is taken to be the definition of inductance. The direction of the voltage drop (plus or minus sign) is such that the inductor resists the change in current.

There's one very intriguing thing about this result. Suppose, for concreteness, that the black box in figure l is a resistor, and that the inductor's energy is decreasing, and being converted into heat in the resistor. The voltage drop across the resistor indicates that it has an electric field across it, which is driving the current. But where is this electric field coming from? There are no charges anywhere that could be creating it! What we've discovered is one special case of a more general principle, the principle of induction: a changing magnetic field creates an electric field, which is in addition to any electric field created by charges. (The reverse is also true: any electric field that changes over time creates a magnetic field.) Induction forms the basis for such technologies as the generator and the transformer, and ultimately it leads to the existence of light, which is a wave pattern in the electric and magnetic fields. These are all topics for chapter 6, but it's truly remarkable that we could come to this conclusion without yet having learned any details about magnetism.

The cartoons in figure m compares electric fields made by charges, 1, to electric fields made by changing magnetic fields, 2-3. In m/1, two physicists are in a room whose ceiling is positively charged and whose floor is negatively charged. The physicist on the bottom throws a positively charged bowling ball into the curved pipe. The physicist at the top uses a radar gun to measure the speed of the ball as it comes out of the pipe. They find that the ball has slowed down by the time it gets to the top. By measuring the change in the ball's kinetic energy, the two physicists are acting just like a voltmeter. They conclude that the top of the tube is at a higher voltage than the bottom of the pipe. A difference in voltage indicates an




m / Electric fields made by charges, 1, and by changing magnetic fields, 2 and 3.

electric field, and this field is clearly being caused by the charges in the floor and ceiling.

In m/2, there are no charges anywhere in the room except for the charged bowling ball. Moving charges make magnetic fields, so there is a magnetic field surrounding the helical pipe while the ball is moving through it. A magnetic field has been created where there was none before, and that field has energy. Where could the energy have come from? It can only have come from the ball itself, so the ball must be losing kinetic energy. The two physicists working together are again acting as a voltmeter, and again they conclude that there is a voltage difference between the top and bottom of the pipe. This indicates an electric field, but this electric field can't have been created by any charges, because there aren't any in the room. This electric field was created by the change in the magnetic field.

The bottom physicist keeps on throwing balls into the pipe, until the pipe is full of balls, m/3, and finally a steady current is established. While the pipe was filling up with balls, the energy in the magnetic field was steadily increasing, and that energy was being stolen from the balls' kinetic energy. But once a steady current is established, the energy in the magnetic field is no longer changing. The balls no longer have to give up energy in order to build up the field, and the physicist at the top finds that the balls are exiting the pipe at full speed again. There is no voltage difference any more. Although there is a current, $\Delta I/\Delta t$ is zero.

### Discussion Questions

**A** What happens when the physicist at the bottom in figure m/3 starts getting tired, and decreases the current?

## A.4 Decay

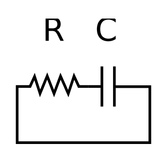
Up until now I've soft-pedaled the fact that by changing the characteristics of an oscillator, it is possible to produce non-oscillatory behavior. For example, imagine taking the mass-on-a-spring system and making the spring weaker and weaker. In the limit of small $k$, it's as though there was no spring whatsoever, and the behavior of the system is that if you kick the mass, it simply starts slowing down. For friction proportional to $v$, as we've been assuming, the result is that the velocity approaches zero, but never actually reaches zero. This is unrealistic for the mechanical oscillator, which will not have vanishing friction at low velocities, but it is quite realistic in the case of an electrical circuit, for which the voltage drop across the resistor really does approach zero as the current approaches zero.

Electrical circuits can exhibit all the same behavior. For simplicity we will analyze only the cases of LRC circuits with $L = 0$ or $C = 0$.

### The RC circuit

We first analyze the RC circuit, n. In reality one would have to "kick" the circuit, for example by briefly inserting a battery, in order to get any interesting behavior. We start with Ohm's law and the equation for the voltage across a capacitor:

$$V_R = IR$$
$$V_C = q/C$$




n / An RC circuit.

The loop rule tells us

$$V_R + V_C = 0 \qquad ,$$

and combining the three equations results in a relationship between $q$ and $I$:

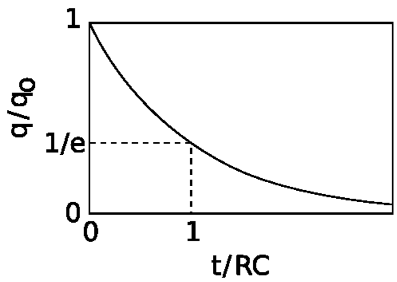
$$I = -\frac{1}{RC}q$$

The negative sign tells us that the current tends to reduce the charge on the capacitor, i.e. to discharge it. It makes sense that the current is proportional to $q$: if $q$ is large, then the attractive forces between the $+q$ and $-q$ charges on the plates of the capacitor are large, and charges will flow more quickly through the resistor in order to reunite. If there was zero charge on the capacitor plates, there would be no reason for current to flow. Since amperes, the unit of current, are the same as coulombs per second, it appears that the quantity $RC$ must have units of seconds, and you can check for yourself that this is correct. $RC$ is therefore referred to as the time constant of the circuit.

How exactly do $I$ and $q$ vary with time? Rewriting $I$ as $\Delta q/\Delta t$, we have

$$\frac{\Delta q}{\Delta t} = -\frac{1}{RC}q \qquad .$$

This equation describes a function $q(t)$ that always gets smaller over time, and whose rate of decrease is big at first, when $q$ is big, but gets smaller and smaller as $q$ approaches zero. As an example of this type of mathematical behavior, we could imagine a man who has 1024 weeds in his backyard, and resolves to pull out half of them every day. On the first day, he pulls out half, and has 512 left. The next day, he pulls out half of the remaining ones, leaving 256. The sequence continues exponentially: 128, 64, 32, 16, 8, 4, 2, 1. Returning to our electrical example, the function $q(t)$ apparently needs to be an exponential, which we can write in the form $ae^{bt}$, where $e = 2.718...$ is the base of natural logarithms. We could have written it with base 2, as in the story of the weeds, rather than base $e$, but the math later on turns out simpler if we use $e$. It doesn't make sense to plug a number that has units into a function like an exponential, so $bt$ must be unitless, and $b$ must therefore have units of inverse seconds. The number $b$ quantifies how fast the exponential decay is. The only physical parameters of the circuit on which $b$ could possibly depend are $R$ and $C$, and the only way to put units of ohms and farads together to make units of inverse seconds is by computing $1/RC$. Well, actually we could use $7/RC$ or $3\pi/RC$, or any other unitless number divided by $RC$, but this is where the use of base $e$ comes in handy: for base $e$, it turns out that the correct unitless constant *is* 1. Thus our solution is

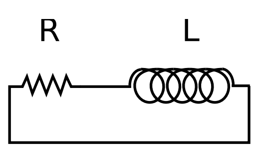



o / Over a time interval $RC$, the charge on the capacitor is reduced by a factor of $e$.

$$q = q_o \exp\left(-\frac{t}{RC}\right) \qquad .$$

The number $RC$, with units of seconds, is called the RC time constant of the circuit, and it tells us how long we have to wait if we want the charge to fall off by a factor of $1/e$.

### The RL circuit




p / An RL circuit.

The RL circuit, p, can be attacked by similar methods, and it can easily be shown that it gives

$$I = I_o \exp\left(-\frac{R}{L}t\right) \qquad .$$

The RL time constant equals $L/R$.

*Death by solenoid; spark plugs* *example 6*

When we suddenly break an RL circuit, what will happen? It might seem that we're faced with a paradox, since we only have two forms of energy, magnetic energy and heat, and if the current stops suddenly, the magnetic field must collapse suddenly. But where does the lost magnetic energy go? It can't go into resistive heating of the resistor, because the circuit has now been broken, and current can't flow!

The way out of this conundrum is to recognize that the open gap in the circuit has a resistance which is large, but not infinite. This large resistance causes the RL time constant $L/R$ to be very small. The current thus continues to flow for a very brief time, and flows straight across the air gap where the circuit has been opened. In other words, there is a spark!

We can determine based on several different lines of reasoning that the voltage drop from one end of the spark to the other must be very large. First, the air's resistance is large, so $V = IR$ requires a large voltage. We can also reason that all the energy in the magnetic field is being dissipated in a short time, so the power dissipated in the spark, $P = IV$, is large, and this requires a large value of $V$. ($I$ isn't large — it is decreasing from its initial value.) Yet a third way to reach the same result is to consider the equation $V_L = \Delta I/\Delta t$: since the time constant is short, the time derivative $\Delta I/\Delta t$ is large.

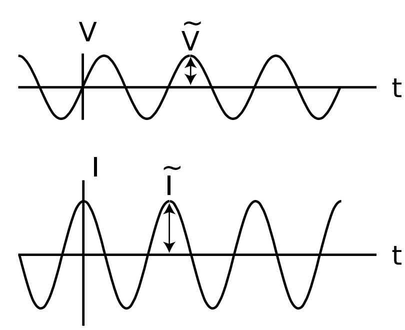
This is exactly how a car's spark plugs work. Another application is to electrical safety: it can be dangerous to break an inductive circuit suddenly, because so much energy is released in a short time. There is also no guarantee that the spark will discharge across the air gap; it might go through your body instead, since your body might have a lower resistance.

### Discussion Questions

**A** A gopher gnaws through one of the wires in the DC lighting system in your front yard, and the lights turn off. At the instant when the circuit becomes open, we can consider the bare ends of the wire to be like the plates of a capacitor, with an air gap (or gopher gap) between them. What kind of capacitance value are we talking about here? What would this tell you about the $RC$ time constant?

## A.5 Impedance

So far we have been thinking in terms of the free oscillations of a circuit. This is like a mechanical oscillator that has been kicked but then left to oscillate on its own without any external force to keep the vibrations from dying out. Suppose an LRC circuit is driven with a sinusoidally varying voltage, such as will occur when a radio tuner is hooked up to a receiving antenna. We know that a current will flow in the circuit, and we know that there will be resonant behavior, but it is not necessarily simple to relate current to voltage in the most general case. Let's start instead with the special cases of LRC circuits consisting of only a resistance, only a capacitance, or only an inductance. We are interested only in the steady-state response.



q / In a capacitor, the current is 90° ahead of the voltage in phase.

The purely resistive case is easy. Ohm's law gives

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

In the purely capacitive case, the relation $V = q/C$ lets us calculate

$$\begin{aligned} I &= \frac{\Delta q}{\Delta t} \\ &= C\frac{\Delta V}{\Delta t} \qquad . \end{aligned}$$

If the voltage varies as, for example, $V(t) = \tilde{V}\sin(\omega t)$, then the current will be $I(t) = \omega C\tilde{V}\cos(\omega t)$, so the maximum current is $\tilde{I} = \omega C\tilde{V}$. By analogy with Ohm's law, we can then write

$$\tilde{I} = \frac{\tilde{V}}{Z_C} \quad ,$$

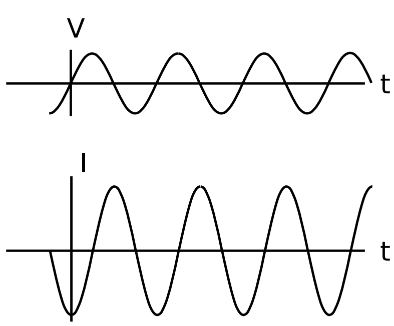
where the quantity

$$Z_C = \frac{1}{\omega C} \quad , \qquad \text{[impedance of a capacitor]}$$

having units of ohms, is called the *impedance* of the capacitor at this frequency. Note that it is only the *maximum* current, $\tilde{I}$, that is proportional to the *maximum* voltage, $\tilde{V}$, so the capacitor is not behaving like a resistor. The maxima of $V$ and $I$ occur at different times, as shown in figure q. It makes sense that the impedance becomes infinite at zero frequency. Zero frequency means that it would take an infinite time before the voltage would change by any amount. In other words, this is like a situation where the capacitor has been connected across the terminals of a battery and been allowed to settle down to a state where there is constant charge on both terminals. Since the electric fields between the plates are constant, there is no energy being added to or taken out of the field. A capacitor that can't exchange energy with any other circuit component is nothing more than a broken (open) circuit.

*self-check C*

Why can't a capacitor have its impedance printed on it along with its capacitance? ▷ Answer, p. 196




r / The current through an inductor lags behind the voltage by a phase angle of 90°.

Similar math gives

$$Z_L = \omega L \qquad \text{[impedance of an inductor]}$$

for an inductor. It makes sense that the inductor has lower impedance at lower frequencies, since at zero frequency there is no change in the magnetic field over time. No energy is added to or released from the magnetic field, so there are no induction effects, and the inductor acts just like a piece of wire with negligible resistance. The term "choke" for an inductor refers to its ability to "choke out" high frequencies.

The phase relationships shown in figures q and r can be remembered using my own mnemonic, "eVIL," which shows that the voltage (V) leads the current (I) in an inductive circuit, while the opposite is true in a capacitive one. A more traditional mnemonic is "ELI the ICE man," which uses the notation E for emf, a concept closely related to voltage.

*Low-pass and high-pass filters* — *example 7*

An LRC circuit only responds to a certain range (band) of frequencies centered around its resonant frequency. As a filter, this is known as a bandpass filter. If you turn down both the bass and the treble on your stereo, you have created a bandpass filter.

To create a high-pass or low-pass filter, we only need to insert a capacitor or inductor, respectively, in series. For instance, a very basic surge protector for a computer could be constructed by inserting an inductor in series with the computer. The desired 60 Hz power from the wall is relatively low in frequency, while the surges that can damage your computer show much more rapid time variation. Even if the surges are not sinusoidal signals, we can think of a rapid "spike" qualitatively as if it was very high in frequency — like a high-frequency sine wave, it changes very rapidly.

Inductors tend to be big, heavy, expensive circuit elements, so a simple surge protector would be more likely to consist of a capacitor in *parallel* with the computer. (In fact one would normally just connect one side of the power circuit to ground via a capacitor.) The capacitor has a very high impedance at the low frequency of the desired 60 Hz signal, so it siphons off very little of the current. But for a high-frequency signal, the capacitor's impedance is very small, and it acts like a zero-impedance, easy path into which the current is diverted.

The main things to be careful about with impedance are that (1) the concept only applies to a circuit that is being driven sinusoidally, (2) the impedance of an inductor or capacitor is frequency-dependent, and (3) impedances in parallel and series don't combine according to the same rules as resistances. It is possible, however, to get get around the third limitation, as discussed in subsection .

### Discussion Question

**A** Figure q on page 179 shows the voltage and current for a capacitor. Sketch the *q*-*t* graph, and use it to give a physical explanation of the phase relationship between the voltage and current. For example, why is the current zero when the voltage is at a maximum or minimum?

**B** Relate the features of the graph in figure r on page 180 to the story told in cartoons in figure m/2-3 on page 176.

## Problems

### Key

- $\checkmark$ A computerized answer check is available online.
- $\int$ A problem that requires calculus.
- $\star$ A difficult problem.

**1** If an FM radio tuner consisting of an LRC circuit contains a 1.0 $\mu$H inductor, what range of capacitances should the variable capacitor be able to provide? $\checkmark$

**2** (a) Show that the equation $V_L = L\,\Delta I/\Delta t$ has the right units.
(b) Verify that $RC$ has units of time.
(c) Verify that $L/R$ has units of time.

**3** Find the energy stored in a capacitor in terms of its capacitance and the voltage difference across it. $\checkmark$

**4** Find the inductance of two identical inductors in parallel.

**5** The wires themselves in a circuit can have resistance, inductance, and capacitance. Would "stray" inductance and capacitance be most important for low-frequency or for high-frequency circuits? For simplicity, assume that the wires act like they're in *series* with an inductor or capacitor.

**6** (a) Find the capacitance of two identical capacitors in series.
(b) Based on this, how would you expect the capacitance of a parallel-plate capacitor to depend on the distance between the plates?

**7** Find the capacitance of the surface of the earth, assuming there is an outer spherical "plate" at infinity. (In reality, this outer plate would just represent some distant part of the universe to which we carried away some of the earth's charge in order to charge up the earth.) $\checkmark$

**8** Starting from the relation $V = L\Delta I/\Delta t$ for the voltage difference across an inductor, show that an inductor has an impedance equal to $L\omega$.

# Appendix 1: Exercises

## Exercise 1A: Electric and Magnetic Forces

Apparatus:

In this exercise, you are going to investigate the forces that can occur among the following objects:

nails

magnets

small bits of paper

specially prepared pieces of scotch tape

To make the specially prepared pieces of tape, take a piece of tape, bend one end over to form a handle that won't stick to your hand, and stick it on a desk. Make a handle on a second piece, and lay it right on top of the first one. Now pull the two pieces off the desk and separate them.

Your goal is to address the following questions experimentally:

1. Do the forces get weaker with distance? Do they have some maximum range? Is there some range at which they abruptly cut off?

2. Can the forces be blocked or shielded against by putting your hand or your calculator in the way? Try this with both electric and magnetic forces, and with both repulsion and attraction.

3. Are the forces among these objects gravitational?

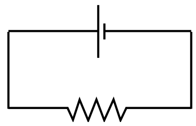
4. Of the many forces that can be observed between different pairs of objects, is there any natural way to classify them into general types of forces?

5. Do the forces obey Newton's third law?

6. Do ordinary materials like wood or paper participate in these forces?

## Exercise 3A: Voltage and Current

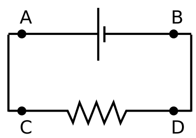
1. How many different currents could you measure in this circuit? Make a prediction, and then try it.



What do you notice? How does this make sense in terms of the roller coaster metaphor introduced in discussion question 3.3A?

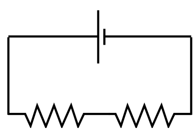
What is being *used up* in the resistor?

2. By connecting probes to these points, how many ways could you measure a voltage? How many of them would be different numbers? Make a prediction, and then do it.




What do you notice? Interpret this using the roller coaster metaphor, and color in parts of the circuit that represent constant voltages.

3. The resistors are unequal. How many *different* voltages and currents can you measure? Make a prediction, and then try it.



What do you notice? Interpret this using the roller coaster metaphor, and color in parts of the circuit that represent constant voltages.

# Exercise 4A: The Loop and Junction Rules

Apparatus:

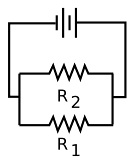
DC power supply

multimeter

resistors

1. The junction rule

Construct a circuit like this one, using the power supply as your voltage source. To make things more interesting, don't use equal resistors. Use nice big resistors (say 100 $k\Omega$ to 1 $M\Omega$) — this will ensure that you don't burn up the resistors, and that the multimeter's small internal resistance when used as an ammeter is negligible in comparison.




Insert your multimeter in the circuit to measure all three currents that you need in order to test the junction rule.

2. The loop rule

Now come up with a circuit to test the loop rule. Since the loop rule is always supposed to be true, it's hard to go wrong here! Make sure you have at least three resistors in a loop, and make sure you hook in the power supply in a way that creates non-zero voltage differences across all the resistors. Measure the voltage differences you need to measure to test the loop rule. Here it is best to use fairly small resistances, so that the multimeter's large internal resistance when used in parallel as a voltmeter will not significantly reduce the resistance of the circuit. Do not use resistances of less than about 100 Ω, however, or you may blow a fuse or burn up a resistor.

## Exercise 4B: Reasoning About Circuits

The questions in this exercise can all be solved using some combination of the following approaches:

a) There is constant voltage throughout any conductor.

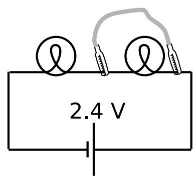
b) Ohm's law can be applied to any *part* of a circuit.

c) Apply the loop rule.

d) Apply the junction rule.

In each case, discuss the question, decide what you think is the right answer, and then try the experiment.

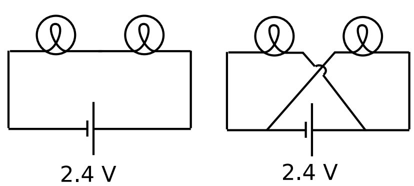
1. A wire is added in parallel with one bulb.




Which reasoning is correct?

- *Each bulb still has 1.2 V across it, so both bulbs are still lit up.*
- *All parts of a wire are at the same voltage, and there is now a wire connection from one side of the right-hand bulb to the other. The right-hand bulb has no voltage difference across it, so it goes out.*

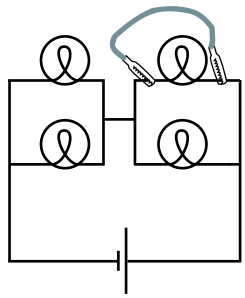
2. The series circuit is changed as shown.




Which reasoning is correct?

- *Each bulb now has its sides connected to the two terminals of the battery, so each now has 2.4 V across it instead of 1.2 V. They get brighter.*
- *Just as in the original circuit, the current goes through one bulb, then the other. It's just that now the current goes in a figure-8 pattern. The bulbs glow the same as before.*

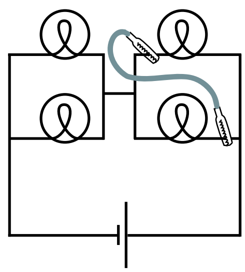
3. A wire is added as shown to the original circuit.



What is wrong with the following reasoning?

*The top right bulb will go out, because its two sides are now connected with wire, so there will be no voltage difference across it. The other three bulbs will not be affected.*

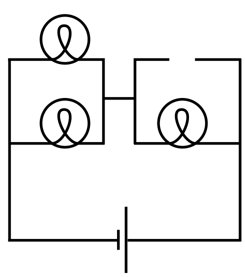
4. A wire is added as shown to the original circuit.



What is wrong with the following reasoning?

*The current flows out of the right side of the battery. When it hits the first junction, some of it will go left and some will keep going up The part that goes up lights the top right bulb. The part that turns left then follows the path of least resistance, going through the new wire instead of the bottom bulb. The top bulb stays lit, the bottom one goes out, and others stay the same.*

5. What happens when one bulb is unscrewed, leaving an air gap?



## Exercise 5A - Field Vectors

Apparatus:

3 solenoids

DC power supply

compass

ruler

cut-off plastic cup

At this point you've studied the gravitational field, **g**, and the electric field, **E**, but not the magnetic field, **B**. However, they all have some of the same mathematical behavior: they act like vectors. Furthermore, magnetic fields are the easiest to manipulate in the lab. Manipulating gravitational fields directly would require futuristic technology capable of moving planet-sized masses around! Playing with electric fields is not as ridiculously difficult, but static electric charges tend to leak off through your body to ground, and static electricity effects are hard to measure numerically. Magnetic fields, on the other hand, are easy to make and control. Any moving charge, i.e. any current, makes a magnetic field.

A practical device for making a strong magnetic field is simply a coil of wire, formally known as a solenoid. The field pattern surrounding the solenoid gets stronger or weaker in proportion to the amount of current passing through the wire.

1. With a single solenoid connected to the power supply and laid with its axis horizontal, use a magnetic compass to explore the field pattern inside and outside it. The compass shows you the field vector's direction, but not its magnitude, at any point you choose. Note that the field the compass experiences is a combination (vector sum) of the solenoid's field and the earth's field.

2. What happens when you bring the compass extremely far away from the solenoid?

What does this tell you about the way the solenoid's field varies with distance?

Thus although the compass doesn't tell you the field vector's magnitude numerically, you can get at least some general feel for how it depends on distance.

3. Make a sea-of-arrows sketch of the magnetic field in the horizontal plane containing the solenoid's axis. The length of each arrow should at least approximately reflect the strength of the magnetic field at that point.

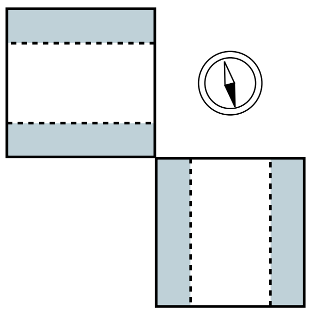
Does the field seem to have sources or sinks?

4. What do you think would happen to your sketch if you reversed the wires?

Try it.

5. Now hook up the two solenoids in parallel. You are going to measure what happens when their two fields combine in the at a certain point in space. As you've seen already, the solenoids' nearby fields are much stronger than the earth's field; so although we now theoretically have three fields involved (the earth's plus the two solenoids'), it will be safe to ignore the earth's field. The basic idea here is to place the solenoids with their axes at some angle to each other, and put the compass at the intersection of their axes, so that it is the same distance from each solenoid. Since the geometry doesn't favor either solenoid, the only factor that would make one solenoid influence the compass more than the other is current. You can use the cut-off plastic cup as a little platform to bring the compass up to the same level as the solenoids' axes.

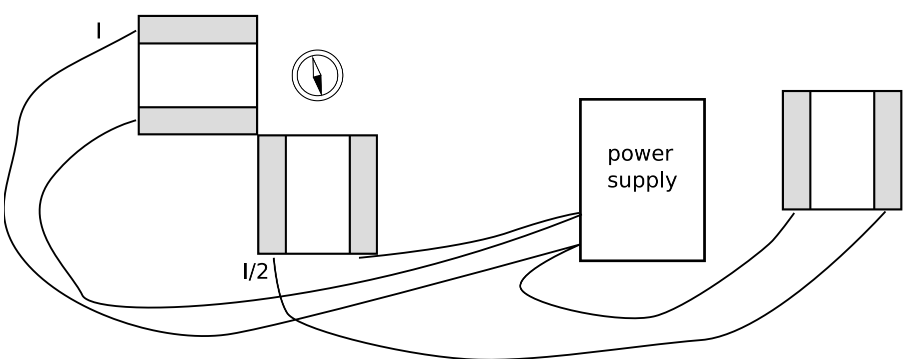
a)What do you think will happen with the solenoids' axes at 90 degrees to each other, and equal currents? Try it. Now represent the vector addition of the two magnetic fields with a diagram. Check your diagram with your instructor to make sure you're on the right track.



b) Now try to make a similar diagram of what would happen if you switched the wires on one of the solenoids.

After predicting what the compass will do, try it and see if you were right.

c)Now suppose you were to go back to the arrangement you had in part a, but you changed one of the currents to half its former value. Make a vector addition diagram, and use trig to predict the angle.




Try it. To cut the current to one of the solenoids in half, an easy and accurate method is simply to put the third solenoid in series with it, and put that third solenoid so far away that its magnetic field doesn't have any significant effect on the compass.

## Exercise 6A - Polarization

Apparatus:

calcite (Iceland spar) crystal

polaroid film

1. Lay the crystal on a piece of paper that has print on it. You will observe a double image. See what happens if you rotate the crystal.

Evidently the crystal does something to the light that passes through it on the way from the page to your eye. One beam of light enters the crystal from underneath, but two emerge from the top; by conservation of energy the energy of the original beam must be shared between them. Consider the following three possible interpretations of what you have observed:

(a) The two new beams differ from each other, and from the original beam, only in energy. Their other properties are the same.

(b) The crystal adds to the light some mysterious new property (not energy), which comes in two flavors, X and Y. Ordinary light doesn't have any of either. One beam that emerges from the crystal has some X added to it, and the other beam has Y.

(c) There is some mysterious new property that is possessed by all light. It comes in two flavors, X and Y, and most ordinary light sources make an equal mixture of type X and type Y light. The original beam is an even mixture of both types, and this mixture is then split up by the crystal into the two purified forms.

In parts 2 and 3 you'll make observations that will allow you to figure out which of these is correct.

2. Now place a polaroid film over the crystal and see what you observe. What happens when you rotate the film in the horizontal plane? Does this observation allow you to rule out any of the three interpretations?

3. Now put the polaroid film under the crystal and try the same thing. Putting together all your observations, which interpretation do you think is correct?

4. Look at an overhead light fixture through the polaroid, and try rotating it. What do you observe? What does this tell you about the light emitted by the lightbulb?

5. Now position yourself with your head under a light fixture and directly over a shiny surface,

such as a glossy tabletop. You'll see the lamp's reflection, and the light coming from the lamp to your eye will have undergone a reflection through roughly a 180-degree angle (i.e. it very nearly reversed its direction). Observe this reflection through the polaroid, and try rotating it. Finally, position yourself so that you are seeing glancing reflections, and try the same thing. Summarize what happens to light with properties X and Y when it is reflected. (This is the principle behind polarizing sunglasses.)

# Appendix 2: Photo Credits

Except as specifically noted below or in a parenthetical credit in the caption of a figure, all the illustrations in this book are under my own copyright, and are copyleft licensed under the same license as the rest of the book.

In some cases it's clear from the date that the figure is public domain, but I don't know the name of the artist or photographer; I would be grateful to anyone who could help me to give proper credit. I have assumed that images that come from U.S. government web pages are copyright-free, since products of federal agencies fall into the public domain. I've included some public-domain paintings; photographic reproductions of them are not copyrightable in the U.S. (Bridgeman Art Library, Ltd. v. Corel Corp., 36 F. Supp. 2d 191, S.D.N.Y. 1999).

When "PSSC Physics" is given as a credit, it indicates that the figure is from the first edition of the textbook entitled Physics, by the Physical Science Study Committee. The early editions of these books never had their copyrights renewed, and are now therefore in the public domain. There is also a blanket permission given in the later PSSC College Physics edition, which states on the copyright page that "The materials taken from the original and second editions and the Advanced Topics of PSSC PHYSICS included in this text will be available to all publishers for use in English after December 31, 1970, and in translations after December 31, 1975."

Credits to Millikan and Gale refer to the textbooks Practical Physics (1920) and Elements of Physics (1927). Both are public domain. (The 1927 version did not have its copyright renewed.) Since is possible that some of the illustrations in the 1927 version had their copyrights renewed and are still under copyright, I have only used them when it was clear that they were originally taken from public domain sources.

In a few cases, I have made use of images under the fair use doctrine. However, I am not a lawyer, and the laws on fair use are vague, so you should not assume that it's legal for you to use these images. In particular, fair use law may give you less leeway than it gives me, because I'm using the images for educational purposes, and giving the book away for free. Likewise, if the photo credit says "courtesy of ...," that means the copyright owner gave me permission to use it, but that doesn't mean you have permission to use it.

**Cover** *Cat neuron:* Courtesy of Richard J. Harris, University of Western Ontario.. **Cover** *Accelerator:* Courtesy of A. Zachau, GSI.. **Contents** *Lightning:* C. Clark/NOAA photo library, public domain. **Contents** *Electric eel:* Wikipedia user Vsion, dual licensed under GFDL and CC-BY. **Contents** *CPU die:* Wikipedia user Matt Gibbs, GFDL licensed. **Contents** *Mars Reconnaissance Orbiter antenna:* NASA, public domain. **??** *Lightning:* C. Clark/NOAA photo library, public domain. **26** *Robert Millikan:* Clark Millikan, 1891, public domain. **31** *J.J. Thomson:* Millikan and Gale, 1920.. **41** *Curies:* Harper's Monthly, 1904. **42** *Becquerel:* Millikan and Gale, 1920. **42** *Becquerel's photographic plate:* public domain. **45** *Rutherford:* public domain. **44** *Nuclear fuel pellets:* US DOE, public domain. **57** *Nuclear power plant:* Wikipedia user Stefan Kuhn, GFDL licensed. **62** *Raft in neutrino detector:* Kamioka Observatory, ICRR (Institute for Cosmic Ray Research), The University of Tokyo. **63** *IceCube:* John Jacobsen/NSF. "This material is based upon work supported by the National Science Foundation under Grant Nos. OPP-9980474 (AMANDA) and OPP-0236449 (IceCube), University of Wisconsin-Madison.". **64** *GAMMASPHERE:* courtesy of C.J. Lister/R.V.F. Janssens. **64** *H bomb test:* public domain product of US DOE, Ivy Mike test. **64** *Fatu Hiva Rainforest:* Wikipedia user Makemake, GFDL licensed. **64** *fusion reactor:* "These images may be used free of charge for educational purposes but please use the acknowledgement 'photograph courtesy of EFDA-JET"'. **64** *Sun:* SOHO (ESA & NASA). **69** *Chernobyl village:* Wikipedia user Slawojarek, GFDL licensed. **69** *Chernobyl map:* CIA Handbook of International Economic Statistics, 1996, public domain. **71** *Crab Nebula:* "ESO Press Photos may be reproduced, if credit is given to the European Southern Observatory.". **77** *Transmission lines:* Wikipedia user Nixdorf, GFDL licensed. **78** *Knifefish:* Courtesy of Greg DeGreef.. **79** *Ampère:* Millikan and Gale, 1920. **83** *Volta:* Millikan and Gale, 1920. **88** *Ohm:* Millikan and Gale, 1920. **90** *Superconducting accelerator segment:* Courtesy of Argonne National Laboratory, managed and operated by the University of Chicago for the U.S. Department of Energy under contract No. W-31-109-ENG-38.. **105** *Printed circuit board:* Bill Bertram, Wikipedia user Pixel8, CC-BY licensed. **107** *CPU die:* Wikipedia user Matt Gibbs, GFDL licensed. **128** *LIGO:* Wikipedia user Umptanum: "This image is copyrighted. The copyright holder has irrevocably released all rights to it.". **135** *Topographical maps:* Flat maps by USGS, public domain; perspective map by Wikipedia user Kbh3rd, GFDL licensed. **143** *Maxwell and his wife:* public domain contemporary photograph. **155** *Hertz:* Public domain contemporary photograph. **149** *Faraday:* Painting by Thomas Phillips, 1842. **??** *Capacitors:*

Wikipedia user de:Benutzer:Honina, GFDL licensed. **??** *Inductors:* Wikipedia user de:Benutzer:Honina, GFDL licensed.

# Appendix 3: Hints and Solutions

## Answers to Self-Checks

### Answers to Self-Checks for Chapter 1

**Page 17, self-check A:** Either type can be involved in either an attraction or a repulsion. A positive charge could be involved in either an attraction (with a negative charge) or a repulsion (with another positive), and a negative could participate in either an attraction (with a positive) or a repulsion (with a negative).

**Page 18, self-check B:** It wouldn't make any difference. The roles of the positive and negative charges in the paper would be reversed, but there would still be a net attraction.

**Page 28, self-check C:** Yes. In U.S. currency, the quantum of money is the penny.

**Page 56, self-check A:** Thomson was accelerating electrons, which are negatively charged. This apparatus is supposed to accelerated atoms with one electron stripped off, which have positive net charge. In both cases, a particle that is between the plates should be attracted by the forward plate and repelled by the plate behind it.

**Page 66, self-check B:** The hydrogen-1 nucleus is simple a proton. The binding energy is the energy required to tear a nucleus apart, but for a nucleus this simple there is nothing to tear apart.

### Answers to Self-Checks for Chapter 3

**Page 92, self-check A:** The large amount of power means a high rate of conversion of the battery's chemical energy into heat. The battery will quickly use up all its energy, i.e., "burn out."

### Answers to Self-Checks for Chapter 5

**Page 129, self-check A:** The reasoning is exactly analogous to that used in example 1 on page 126 to derive an equation for the gravitational field of the earth. The field is $F/q_t = (kQq_t/r^2)/q_t = kQ/r^2$.

**Page 134, self-check B:**

$$
\begin{aligned}
E_x &= -\frac{\mathrm{d}V}{\mathrm{d}x} \\
&= -\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{kQ}{r}\right) \\
&= \frac{kQ}{r^2}
\end{aligned}
$$

**Page 136, self-check C:** (a) The voltage (height) increases as you move to the east or north. If we let the positive $x$ direction be east, and choose positive $y$ to be north, then $\mathrm{d}V/\mathrm{d}x$ and $\mathrm{d}V/\mathrm{d}y$ are both positive. This means that $E_x$ and $E_y$ are both negative, which makes sense, since the water is flowing in the negative $x$ and $y$ directions (south and west).
(b) The electric fields are all pointing away from the higher ground. If this was an electrical map, there would have to be a large concentration of charge all along the top of the ridge, and especially at the mountain peak near the south end.

### Answers to Self-Checks for Chapter 6

**Page 152, self-check A:** An induced electric field can only be created by a *changing* magnetic field. Nothing is changing if your car is just sitting there. A point on the coil won't experience a changing magnetic field unless the coil is already spinning, i.e., the engine has already turned over.

### Answers to Self-Checks for Chapter A

**Page 171, self-check A:** Yes. The mass has the same kinetic energy regardless of which direction it's moving. Friction coverts mechanical energy into heat at the same rate whether the mass is sliding to the right or to the left. The spring has an equilibrium length, and energy can be stored in it either by compressing it ($x < 0$) or stretching it ($x > 0$).

**Page 171, self-check B:** Velocity, $v$, is the rate of change of position, $x$, with respect to time. This is exactly analogous to $I = \Delta q/\Delta t$.

**Page 180, self-check C:** The impedance depends on the frequency at which the capacitor is being driven. It isn't just a single value for a particular capacitor.

## Solutions to Selected Homework Problems

### Solutions for Chapter 2

**Page 75, problem 6:** (a) In the reaction $\mathrm{p}+\mathrm{e}^- \rightarrow \mathrm{n}+\nu$, the charges on the left are $e+(-e) = 0$, and both charges on the right are zero. (b) The neutrino has negligible mass. The masses on the left add up to less than the mass of the neutrino on the right, so energy would be required from an external source in order to make this reaction happen.

### Solutions for Chapter 3

**Page 104, problem 12:** $\Delta$t= Dq/ I $= e/I = 0.160\ \mu$s.

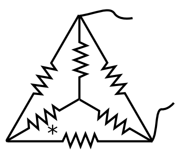
**Page 104, problem 13:** (a) The change in PE is $e\Delta\mathrm{V}$, so the KE gained is $(1/2)mv^2 = eV$. Solving for $v$ and plugging in numbers, we get $5.9 \times 10^7$ m/s. This is about 20% of the speed of light. (Since it's not that close to the speed of light, we'll get a reasonably accurate answer without taking into account Einstein's theory of relativity.)

**Page 105, problem 16:** It's much more practical to measure voltage differences. To measure a current, you have to break the circuit somewhere and insert the meter there, but it's not possible to disconnect the circuits sealed inside the board.

## Solutions for Chapter 4

**Page 120, problem 11:** In series, they give 11 k$\Omega$. In parallel, they give $(1/1\ \mathrm{k}\Omega + 1/10\ \mathrm{k}\Omega)^{-1} = 0.9\ \mathrm{k}\Omega$.

**Page 121, problem 12:** The actual shape is irrelevant; all we care about it what's connected to what. Therefore, we can draw the circuit flattened into a plane. Every vertex of the tetrahedron is adjacent to every other vertex, so any two vertices to which we connect will give the same resistance. Picking two arbitrarily, we have this:




This is unfortunately a circuit that cannot be converted into parallel and series parts, and that's what makes this a hard problem! However, we can recognize that by symmetry, there is zero current in the resistor marked with an asterisk. Eliminating this one, we recognize the whole arrangement as a triple parallel circuit consisting of resistances $R$, $2R$, and $2R$. The resulting resistance is $R/2$.

## Solutions for Chapter 5

**Page 141, problem 9:** Proceeding as suggested in the hint, we form concentric rings, each one extending from radius $b$ to radius $b + \mathrm{d}b$.The area of such a ring equals its circumference multipled by $\mathrm{d}b$, which is $(2\pi b)\mathrm{d}b$. Its charge is thus $2\pi\sigma b\mathrm{d}b$. Plugging this in to the expression from problem 8 gives a contribution to the field $\mathrm{d}E = 2\pi\sigma bka(a^2 + b^2)^{-3/2}\mathrm{d}b$. The total field is found by integrating this expression. The relevant integral can be found in a table.

$$
\begin{aligned}
E &= \int_0^\infty \mathrm{d}E = 2\pi\sigma bka(a^2+b^2)^{-3/2}\mathrm{d}b \\
&= 2\pi\sigma ka \int_0^\infty b(a^2+b^2)^{-3/2}\mathrm{d}b \\
&= 2\pi\sigma ka \left[-\left(a^2+b^2\right)^{-1/2}\right]_{b=0}^{\infty} \\
&= 2\pi\sigma k
\end{aligned}
$$

**Page 141, problem 11:** Let the square's sides be of length $a$. The field at the center is the vector sum of the fields that would have been produced individually by the three charges. Each of these individual fields is $kq/r^2$, where $r_1 = a/\sqrt{2}$ for the two charges $q_1$, and $r_2 = a/2$ for $q_2$. Vector addition can be done by adding components. Let $x$ be horizontal and $y$ vertical. The $y$

components cancel by symmetry. The sum of the $x$ components is

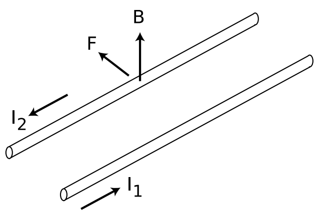
$$E_x = \frac{kq_1}{r_1^2}\cos 45^\circ + \frac{kq_1}{r_1^2}\cos 45^\circ - \frac{kq_2}{r_2^2} \qquad .$$

Substituting $\cos 45^\circ = 1/\sqrt{2}$ and setting this whole expression equal to zero, we find $q_2/q_1 = 1/\sqrt{2}$.

## Solutions for Chapter 6

**Page 164, problem 13:** (a) Current means how much charge passes by a given point per unit time. During a time interval $\Delta$t, all the charge carriers in a certain region behind the point will pass by. This region has length $v\Delta t$ and cross-sectional area $A$, so its volume is $Av\Delta$t, and the amount of charge in it is $Avnq\Delta$t. To find the current, we divide this amount of charge by $\Delta$t, giving $I = Avnq$. (b) A segment of the wire of length $L$ has a force $QvB$ acting on it, where $Q = ALnq$ is the total charge of the moving charge carriers in that part of the wire. The force per unit length is $ALnqvB/L = AnqvB$. (c) Dividing the two results gives $F/L = IB$.

**Page 165, problem 14:** (a) The figure shows the case where the currents are in opposite directions.




The field vector shown is one made by wire 1, which causes an effect on wire 2. It points up because wire 1's field pattern is clockwise as view from along the direction of current $I_1$. For simplicity, let's assume that the current $I_2$ is made by positively charged particles moving in the direction of the current. (You can check that the final result would be the same if they were negatively charged, as would actually be the case in a metal wire.) The force on one of these positively charged particles in wire 2 is supposed to have a direction such that when you sight along it, the $B$ vector is clockwise from the $v$ vector. This can only be accomplished if the force on the particle in wire 2 is in the direction shown. Wire 2 is repelled by wire 1.

To verify that wire 1 is also repelled by wire 2, we can either go through the same type of argument again, or we can simply apply Newton's third law.

Simialar arguments show that the force is attractive if the currents are in the same direction.

(b) The force on wire 2 is $F/L = I_{2B}$, where $B = \mu_o I_1/2\pi r$ is the field made by wire 1 and $r$ is the distance between the wires. The result is

$$F/L = \mu_o I_1 I_2 / 2\pi r \qquad .$$

**Page 166, problem 19:** (a) Based on our knowledge of the field pattern of a current-carrying loop, we know that the magnetic field must be either into or out of the page. This makes sense, since that would mean the field is always perpendicular to the plane of the electrons' motion; if it was in their plane of motion, then the angle between the $v$ and $B$ vectors would be changing all the time, but we see no evidence of such behavior. With the field turned on, the force vector is apparently toward the center of the circle. Let's analyze the force at the moment when the electrons have started moving, which is at the right side of the circle. The force is to the left. Since the electrons are negatively charged particles, we know that if we sight along the force vector, the $B$ vector must be counterclockwise from the $v$ vector. The magnetic field must be out of the page. (b) Looking at figure h on page 147, we can tell that the current in the coils must be counterclockwise as viewed from the perspective of the camera. (c) Electrons are negatively charged, so to produce a counterclockwise current, the electrons in the coils must be going clockwise. (d) The current in the coils is keep the electrons in the beam from going straight, i.e. the force is a repulsion. This makes sense by comparison with problem 14: the coil currents and vacuum tube currents are counterrotating, which causes a repulsion.

**Page 166, problem 20:** Yes. For example, the force vanishes if the particle's velocity is parallel to the field, so if the beam had been launched parallel to the field, it would have gone in a straight line rather than a circle. In general, any component of the velocity vector that is out of the plane perpendicular to the field will remain constant, so the motion can be helical.

**Page 166, problem 22:** The trick is to imagine putting together two identical solenoids to make one double-length solenoid. The field of the doubled solenoid is given by the vector sum of the two solenoids' individual fields. At points on the axis, symmetry guarantees that the individual fields lie along the axis, and similarly for the total field. At the center of one of the mouths, we thus have two parallel field vectors of equal strength, whose sum equals the interior field. But the interior field of the doubled solenoid is the same as that of the individual ones, since the equation for the field only depends on the number of turns per unit length. Therefore the field at the center of a solenoid's mouth equals exactly half the interior field.

# Index

# Useful Data

## Metric Prefixes

| | | |
|---|---|---|
| M- | mega- | $10^6$ |
| k- | kilo- | $10^3$ |
| m- | milli- | $10^{-3}$ |
| $\mu$- (Greek mu) | micro- | $10^{-6}$ |
| n- | nano- | $10^{-9}$ |
| p- | pico- | $10^{-12}$ |
| f- | femto- | $10^{-15}$ |

(Centi-, $10^{-2}$, is used only in the centimeter.)

## Notation and Units

| **quantity** | **unit** | **symbol** |
|---|---|---|
| distance | meter, m | $x, \Delta x$ |
| time | second, s | $t, \Delta t$ |
| mass | kilogram, kg | $m$ |
| density | $kg/m^3$ | $\rho$ |
| velocity | m/s | $\mathbf{v}$ |
| acceleration | $m/s^2$ | $\mathbf{a}$ |
| force | $N = kg{\cdot}m/s^2$ | $\mathbf{F}$ |
| pressure | $Pa{=}1\ N/m^2$ | $P$ |
| energy | $J = kg{\cdot}m^2/s^2$ | $E$ |
| power | W = 1 J/s | $P$ |
| momentum | kg·m/s | $\mathbf{p}$ |
| period | s | $T$ |
| wavelength | m | $\lambda$ |
| frequency | $s^{-1}$ or Hz | $f$ |
| charge | coulomb, C | $q$ |
| voltage | volt, 1 V = 1 J/C | $V$ |
| current | ampere, 1 A = 1 C/s | $I$ |
| resistance | ohm, 1 Ω = 1 V/A | $R$ |
| capacitance | farad, 1 F = 1 C/V | $C$ |
| inductance | henry, 1 H = 1 V·s/A | $L$ |
| electric field | V/m or N/C | $E$ |
| magnetic field | tesla, 1 T = 1 N·s/C·m | $B$ |

## Fundamental Constants

| | |
|---|---|
| gravitational constant | $G = 6.67 \times 10^{-11}\ N{\cdot}m^2/kg^2$ |
| Coulomb constant | $k = 8.99 \times 10^9\ N{\cdot}m^2/C^2$ |
| quantum of charge | $e = 1.60 \times 10^{-19}$ C |
| speed of light | $c = 3.00 \times 10^8$ m/s |

## Conversions

Nonmetric units in terms of metric ones:

| | |
|---|---|
| 1 inch | = 25.4 mm (by definition) |
| 1 pound-force | = 4.5 newtons of force |
| (1 kg) $\cdot g$ | = 2.2 pounds-force |
| 1 scientific calorie | = 4.18 J |
| 1 kcal | $= 4.18 \times 10^3$ J |
| 1 gallon | $= 3.78 \times 10^3\ cm^3$ |
| 1 horsepower | = 746 W |

When speaking of food energy, the word "Calorie" is used to mean 1 kcal, i.e., 1000 calories. In writing, the capital C may be used to indicate 1 Calorie=1000 calories.

Relationships among U.S. units:

| | |
|---|---|
| 1 foot (ft) | = 12 inches |
| 1 yard (yd) | = 3 feet |
| 1 mile (mi) | = 5280 feet |

## Earth, Moon, and Sun

| **body** | **mass (kg)** | **radius (km)** | **radius of orbit (km)** |
|---|---|---|---|
| earth | $5.97 \times 10^{24}$ | $6.4 \times 10^3$ | $1.49 \times 10^8$ |
| moon | $7.35 \times 10^{22}$ | $1.7 \times 10^3$ | $3.84 \times 10^5$ |
| sun | $1.99 \times 10^{30}$ | $7.0 \times 10^5$ | — |

## Subatomic Particles

| **particle** | **mass (kg)** | **radius (fm)** |
|---|---|---|
| electron | $9.109 \times 10^{-31}$ | $\lesssim 0.01$ |
| proton | $1.673 \times 10^{-27}$ | $\sim 1.1$ |
| neutron | $1.675 \times 10^{-27}$ | $\sim 1.1$ |

The radii of protons and neutrons can only be given approximately, since they have fuzzy surfaces. For comparison, a typical atom is about a million fm in radius.