Phys1140: Introductory Physics II: Part 1

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 04/15/2025

TABLE OF CONTENTS

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Charges and Conductors

1.1: Prelude to Electric Charges and Fields

Back when we were studying Newton's laws, we identified several physical phenomena as forces. We did so based on the effect they had on a physical object: Specifically, they caused the object to accelerate. Later, when we studied impulse and momentum, we expanded this idea to identify a force as any physical phenomenon that changed the momentum of an object. In either case, the result is the same: We recognize a force by the effect that it has on an object.

Figure 1.1.1: Electric charges exist all around us. They can cause objects to be repelled from each other or to be attracted to each other. (credit: modification of work by Sean McGrath)

In Gravitation, we examined the force of gravity, which acts on all objects with mass. In this chapter, we begin the study of the electric force, which acts on all objects with a property called charge. The electric force is much stronger than gravity (in most systems where both appear), but it can be a force of attraction or a force of repulsion, which leads to very different effects on objects. The electric force helps keep atoms together, so it is of fundamental importance in matter. But it also governs most everyday interactions we deal with, from chemical interactions to biological processes.

1.2: Electric Charge

Learning Objectives

By the end of this section, you will be able to:

- Describe the concept of electric charge
- Explain qualitatively the force electric charge creates

You are certainly familiar with electronic devices that you activate with the click of a switch, from computers to cell phones to television. And you have certainly seen electricity in a flash of lightning during a heavy thunderstorm. But you have also most likely experienced electrical effects in other ways, maybe without realizing that an electric force was involved. Let's take a look at some of these activities and see what we can learn from them about electric charges and forces.

Discoveries

You have probably experienced the phenomenon of **static electricity**: When you first take clothes out of a dryer, many (not all) of them tend to stick together; for some fabrics, they can be very difficult to separate. Another example occurs if you take a woolen sweater off quickly—you can feel (and hear) the static electricity pulling on your clothes, and perhaps even your hair. If you comb your hair on a dry day and then put the comb close to a thin stream of water coming out of a faucet, you will find that the water stream bends toward (is attracted to) the comb (Figure $1.2.1$).

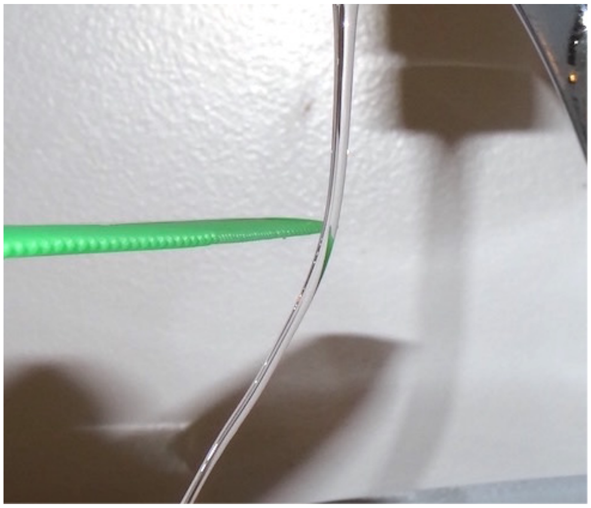


Figure $1.2.1$: An electrically charged comb attracts a stream of water from a distance. Note that the water is not touching the comb. (credit: Jane Whitney)

Suppose you bring the comb close to some small strips of paper; the strips of paper are attracted to the comb and even cling to it (Figure $1.2.2$). In the kitchen, quickly pull a length of plastic cling wrap off the roll; it will tend to cling to most any nonmetallic material (such as plastic, glass, or food). If you rub a balloon on a wall for a few seconds, it will stick to the wall. Probably the most annoying effect of static electricity is getting shocked by a doorknob (or a friend) after shuffling your feet on some types of carpeting.

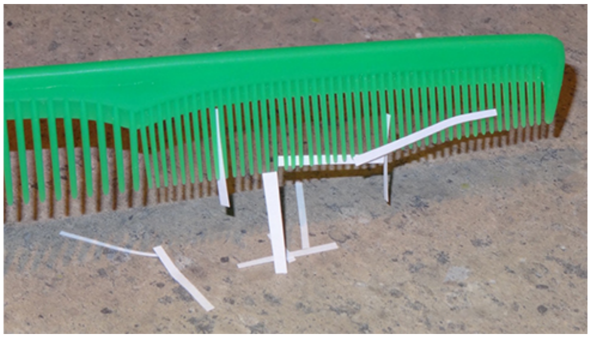


Figure 1.2.2: After being used to comb hair, this comb attracts small strips of paper from a distance, without physical contact. Investigation of this behavior helped lead to the concept of the electric force.

Many of these phenomena have been known for centuries. The ancient Greek philosopher Thales of Miletus (624–546 BCE) recorded that when amber (a hard, translucent, fossilized resin from extinct trees) was vigorously rubbed with a piece of fur, a force was created that caused the fur and the amber to be attracted to each other (Figure 1.2.3). Additionally, he found that the rubbed amber would not only attract the fur, and the fur attract the amber, but they both could affect other (nonmetallic) objects, even if not in contact with those objects (Figure 1.2.4).

Figure 1.2.3: Borneo amber is mined in Sabah, Malaysia, from shale-sandstone-mudstone veins. When a piece of amber is rubbed with a piece of fur, the amber gains more electrons, giving it a net negative charge. At the same time, the fur, having lost electrons, becomes positively charged. (credit: “Sebakoamber”/Wikimedia Commons)

The English physicist William Gilbert (1544–1603) also studied this attractive force, using various substances. He worked with amber, and, in addition, he experimented with rock crystal and various precious and semi-precious gemstones. He also experimented with several metals. He found that the metals never exhibited this force, whereas the minerals did. Moreover, although an electrified amber rod would attract a piece of fur, it would repel another electrified amber rod; similarly, two electrified pieces of fur would repel each other.

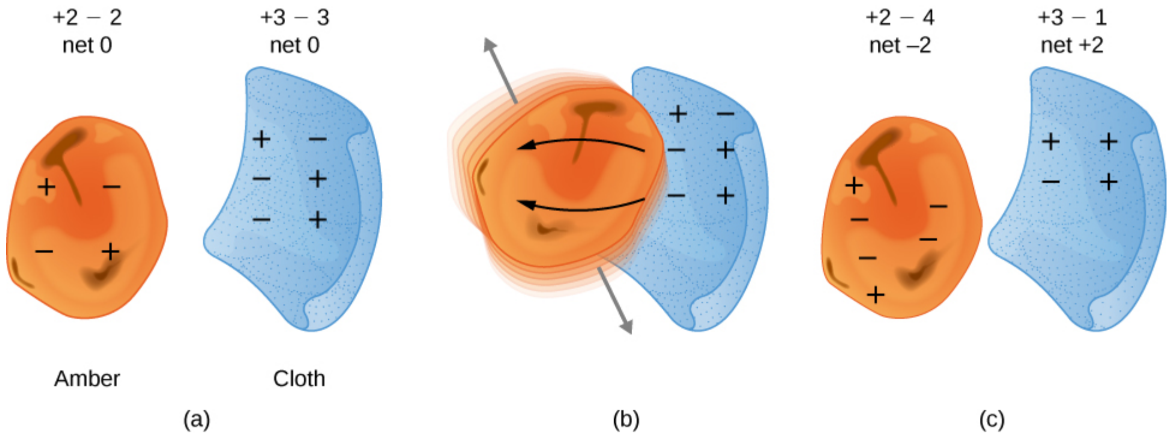


Figure 1.2.4: When materials are rubbed together, charges can be separated, particularly if one material has a greater affinity for electrons than another. (a) Both the amber and cloth are originally neutral, with equal positive and negative charges. Only a tiny fraction of the charges are involved, and only a few of them are shown here. (b) When rubbed together, some negative charge is transferred to the amber, leaving the cloth with a net positive charge. (c) When separated, the amber and cloth now have net charges, but the absolute value of the net positive and negative charges will be equal.

This suggested there were two types of an electric property; this property eventually came to be called **electric charge**. The difference between the two types of electric charge is in the directions of the electric forces that each type of charge causes: These forces are repulsive when the same type of charge exists on two interacting objects and attractive when the charges are of opposite types. The SI unit of electric charge is the **coulomb** (C), after the French physicist Charles Augustine de Coulomb (1736–1806).

The most peculiar aspect of this new force is that it does not require physical contact between the two objects in order to cause an acceleration. This is an example of a so-called "long-range" force. (Or, as James Clerk Maxwell later phrased it, "action at a distance.") With the exception of gravity, all other forces we have discussed so far act only when the two interacting objects actually touch.

The American physicist and statesman Benjamin Franklin found that he could concentrate charge in a "Leyden jar," which was essentially a glass jar with two sheets of metal foil, one inside and one outside, with the glass between them (Figure 1.2.4). This created a large electric force between the two foil sheets.

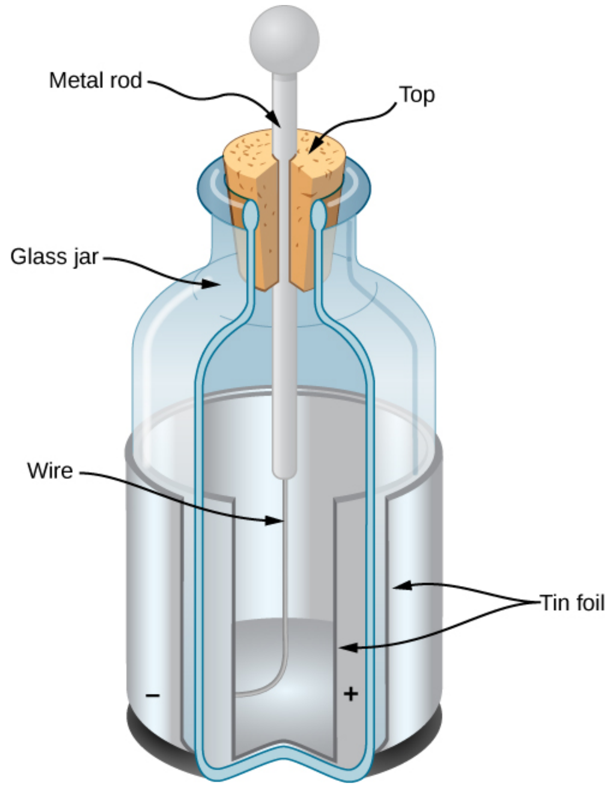


Figure 1.2.4: A Leyden jar (an early version of what is now called a capacitor) allowed experimenters to store large amounts of electric charge. Benjamin Franklin used such a jar to demonstrate that lightning behaved exactly like the electricity he got from the equipment in his laboratory.

Franklin pointed out that the observed behavior could be explained by supposing that one of the two types of charge remained motionless, while the other type of charge flowed from one piece of foil to the other. He further suggested that an excess of what he called this “electrical fluid” be called “positive electricity” and the deficiency of it be called “negative electricity.” His suggestion, with some minor modifications, is the model we use today. (With the experiments that he was able to do, this was a pure guess; he had no way of actually determining the sign of the moving charge. Unfortunately, he guessed wrong; we now know that the charges that flow are the ones Franklin labeled negative, and the positive charges remain largely motionless. Fortunately, as we’ll see, it makes no practical or theoretical difference which choice we make, as long as we stay consistent with our choice.)

Let’s list the specific observations that we have of this **electric force**:

- The force acts without physical contact between the two objects.
- The force can be either attractive or repulsive: If two interacting objects carry the same sign of charge, the force is repulsive; if the charges are of opposite sign, the force is attractive. These interactions are referred to as **electrostatic repulsion** and **electrostatic attraction**, respectively.
- Not all objects are affected by this force.
- The magnitude of the force decreases (rapidly) with increasing separation distance between the objects.

To be more precise, we find experimentally that the magnitude of the force decreases as the square of the distance between the two interacting objects increases. Thus, for example, when the distance between two interacting objects is doubled, the force between them decreases to one fourth what it was in the original system. We can also observe that the surroundings of the charged objects affect the magnitude of the force. However, we will explore this issue in a later chapter.

Properties of Electric Charge

In addition to the existence of two types of charge, several other properties of charge have been discovered.

- **Charge is quantized.** This means that electric charge comes in discrete amounts, and there is a smallest possible amount of charge that an object can have. In the SI system, this smallest amount is $e \equiv 1.602 \times 10^{-19}\ \mathrm{C}$. No free particle can have less charge than this, and, therefore, the charge on any object—the charge on all objects—must be an integer multiple of this amount. All macroscopic, charged objects have charge because electrons have either been added or taken away from them, resulting in a net charge.

- **The magnitude of the charge is independent of the type.** Phrased another way, the smallest possible positive charge (to four significant figures) is $+1.602 \times 10^{-19}$ C, and the smallest possible negative charge is -1.602×10^{-19}; these values are exactly equal. This is simply how the laws of physics in our universe turned out.
- **Charge is conserved.** Charge can neither be created nor destroyed; it can only be transferred from place to place, from one object to another. Frequently, we speak of two charges "canceling"; this is verbal shorthand. It means that if two objects that have equal and opposite charges are physically close to each other, then the (oppositely directed) forces they apply on some other charged object cancel, for a net force of zero. It is important that you understand that the charges on the objects by no means disappear, however. The net charge of the universe is constant.
- **Charge is conserved in closed systems.** In principle, if a negative charge disappeared from your lab bench and reappeared on the Moon, conservation of charge would still hold. However, this never happens. If the total charge you have in your local system on your lab bench is changing, there will be a measurable flow of charge into or out of the system. Again, charges can and do move around, and their effects can and do cancel, but the net charge in your local environment (if closed) is conserved. The last two items are both referred to as the **law of conservation of charge**.

The Source of Charges: The Structure of the Atom

Once it became clear that all matter was composed of particles that came to be called atoms, it also quickly became clear that the constituents of the atom included both positively charged particles and negatively charged particles. The next question was, what are the physical properties of those electrically charged particles?

The negatively charged particle was the first one to be discovered. In 1897, the English physicist J. J. Thomson was studying what was then known as *cathode rays*. Some years before, the English physicist William Crookes had shown that these "rays" were negatively charged, but his experiments were unable to tell any more than that. (The fact that they carried a negative electric charge was strong evidence that these were not rays at all, but particles.) Thomson prepared a pure beam of these particles and sent them through crossed electric and magnetic fields, and adjusted the various field strengths until the net deflection of the beam was zero. With this experiment, he was able to determine the charge-to-mass ratio of the particle. This ratio showed that the mass of the particle was much smaller than that of any other previously known particle—1837 times smaller, in fact. Eventually, this particle came to be called the **electron**.

Since the atom as a whole is electrically neutral, the next question was to determine how the positive and negative charges are distributed within the atom. Thomson himself imagined that his electrons were embedded within a sort of positively charged paste, smeared out throughout the volume of the atom. However, in 1908, the New Zealand physicist Ernest Rutherford showed that the positive charges of the atom existed within a tiny core—called a nucleus—that took up only a very tiny fraction of the overall volume of the atom, but held over 99% of the mass (see Linear Momentum and Collisions.) In addition, he showed that the negatively charged electrons perpetually orbited about this nucleus, forming a sort of electrically charged cloud that surrounds the nucleus (Figure 1.2.5). Rutherford concluded that the nucleus was constructed of small, massive particles that he named **protons**.

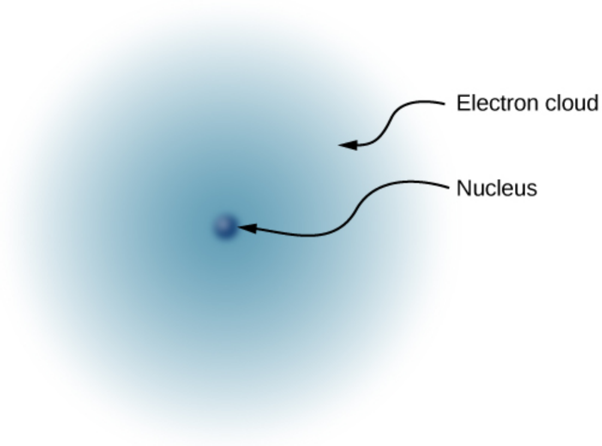


Figure $1.2.5$: This simplified model of a hydrogen atom shows a positively charged nucleus (consisting, in the case of hydrogen, of a single proton), surrounded by an electron "cloud." The charge of the electron cloud is equal (and opposite in sign) to the charge of the nucleus, but the electron does not have a definite location in space; hence, its representation here is as a cloud. Normal macroscopic amounts of matter contain immense numbers of atoms and molecules, and, hence, even greater numbers of individual negative and positive charges.

Since it was known that different atoms have different masses, and that ordinarily atoms are electrically neutral, it was natural to suppose that different atoms have different numbers of protons in their nucleus, with an equal number of negatively charged electrons orbiting about the positively charged nucleus, thus making the atoms overall electrically neutral. However, it was soon discovered that although the lightest atom, hydrogen, did indeed have a single proton as its nucleus, the next heaviest atom—helium—has twice the number of protons (two), but *four* times the mass of hydrogen.

This mystery was resolved in 1932 by the English physicist James Chadwick, with the discovery of the **neutron**. The neutron is, essentially, an electrically neutral twin of the proton, with no electric charge, but (nearly) identical mass to the proton. The helium nucleus therefore has two neutrons along with its two protons. (Later experiments were to show that although the neutron is electrically neutral overall, it does have an internal charge *structure*. Furthermore, although the masses of the neutron and the proton are *nearly* equal, they aren't exactly equal: The neutron's mass is very slightly larger than the mass of the proton. That slight mass excess turned out to be of great importance. That, however, is a story that will have to wait until our study of modern physics in Nuclear Physics.)

Thus, in 1932, the picture of the atom was of a small, massive nucleus constructed of a combination of protons and neutrons, surrounded by a collection of electrons whose combined motion formed a sort of negatively charged "cloud" around the nucleus (Figure $1.2.6$). In an electrically neutral atom, the total negative charge of the collection of electrons is equal to the total positive charge in the nucleus. The very low-mass electrons can be more or less easily removed or added to an atom, changing the net charge on the atom (though without changing its type). An atom that has had the charge altered in this way is called an **ion**. Positive ions have had electrons removed, whereas negative ions have had excess electrons added. We also use this term to describe molecules that are not electrically neutral.

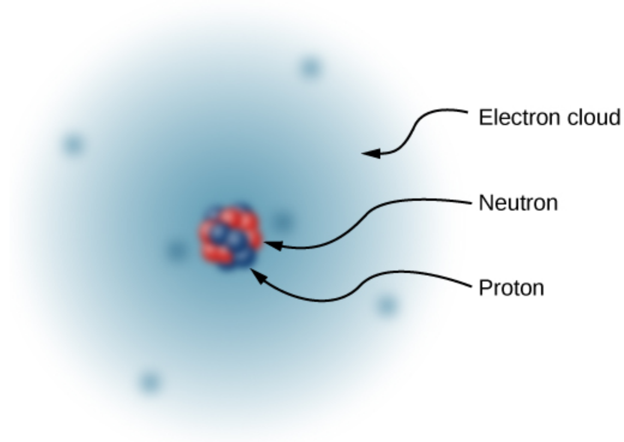


Figure 1.2.6: The nucleus of a carbon atom is composed of six protons and six neutrons. As in hydrogen, the surrounding six electrons do not have definite locations and so can be considered to be a sort of cloud surrounding the nucleus.

The story of the atom does not stop there, however. In the latter part of the twentieth century, many more subatomic particles were discovered in the nucleus of the atom: pions, neutrinos, and quarks, among others. With the exception of the photon, none of these particles are directly relevant to the study of electromagnetism, so we defer further discussion of them until the chapter on particle physics (Particle Physics and Cosmology).

A Note on Terminology

As noted previously, electric charge is a property that an object can have. This is similar to how an object can have a property that we call mass, a property that we call density, a property that we call temperature, and so on. Technically, we should always say something like, "Suppose we have a particle that carries a charge of μC." However, it is very common to say instead, "Suppose we have a μC charge." Similarly, we often say something like, "Six charges are located at the vertices of a regular hexagon." A charge is not a particle; rather, it is a *property* of a particle. Nevertheless, this terminology is extremely common (and is frequently used in this book, as it is everywhere else). So, keep in the back of your mind what we really mean when we refer to a "charge."

1.3: Conductors, Insulators, and Charging by Induction

Learning Objectives

By the end of this section, you will be able to:

- Explain what a conductor is
- Explain what an insulator is
- List the differences and similarities between conductors and insulators
- Describe the process of charging by induction

In the preceding section, we said that scientists were able to create electric charge only on nonmetallic materials and never on metals. To understand why this is the case, you have to understand more about the nature and structure of atoms. In this section, we discuss how and why electric charges do—or do not—move through materials (Figure $1.3.1$). A more complete description is given in a later chapter.

Figure $1.3.1$: This power adapter uses metal wires and connectors to conduct electricity from the wall socket to a laptop computer. The conducting wires allow electrons to move freely through the cables, which are shielded by rubber and plastic. These materials act as insulators that don't allow electric charge to escape outward. (credit: modification of work by "Evan-Amos"/Wikimedia Commons)

Conductors and Insulators

As discussed in the previous section, electrons surround the tiny nucleus in the form of a (comparatively) vast cloud of negative charge. However, this cloud does have a definite structure to it. Let's consider an atom of the most commonly used conductor, copper.

For reasons that will become clear in Atomic Structure, there is an outermost electron that is only loosely bound to the atom's nucleus. It can be easily dislodged; it then moves to a neighboring atom. In a large mass of copper atoms (such as a copper wire or a sheet of copper), these vast numbers of outermost electrons (one per atom) wander from atom to atom, and are the electrons that do the moving when electricity flows. These wandering, or "free," electrons are called **conduction electrons**, and copper is therefore an excellent **conductor** (of electric charge). All conducting elements have a similar arrangement of their electrons, with one or two conduction electrons. This includes most metals.

Insulators, in contrast, are made from materials that lack conduction electrons; charge flows only with great difficulty, if at all. Even if excess charge is added to an insulating material, it cannot move, remaining indefinitely in place. This is why insulating materials exhibit the electrical attraction and repulsion forces described earlier, whereas conductors do not; any excess charge placed on a conductor would instantly flow away (due to mutual repulsion from existing charges), leaving no excess charge around to create forces. Charge cannot flow along or through an **insulator**, so its electric forces remain for long periods of time. (Charge will dissipate from an insulator, given enough time.) As it happens, amber, fur, and most semi-precious gems are insulators, as are materials like wood, glass, and plastic.

Charging by Induction

Let's examine in more detail what happens in a conductor when an electrically charged object is brought close to it. As mentioned, the conduction electrons in the conductor are able to move with nearly complete freedom. As a result, when a charged insulator (such as a positively charged glass rod) is brought close to the conductor, the (total) charge on the insulator exerts an electric force on the conduction electrons. Since the rod is positively charged, the conduction electrons (which themselves are negatively charged) are attracted, flowing toward the insulator to the near side of the conductor (Figure 1.3.2).

Now, the conductor is still overall electrically neutral; the conduction electrons have changed position, but they are still in the conducting material. However, the conductor now has a charge *distribution*; the near end (the portion of the conductor closest to the insulator) now has more negative charge than positive charge, and the reverse is true of the end farthest from the insulator. The relocation of negative charges to the near side of the conductor results in an overall positive charge in the part of the conductor farthest from the insulator. We have thus created an electric charge distribution where one did not exist before. This process is referred to as *inducing polarization*—in this case, polarizing the conductor. The resulting separation of positive and negative charge is called **polarization**, and a material, or even a molecule, that exhibits polarization is said to be polarized. A similar situation occurs with a negatively charged insulator, but the resulting polarization is in the opposite direction.

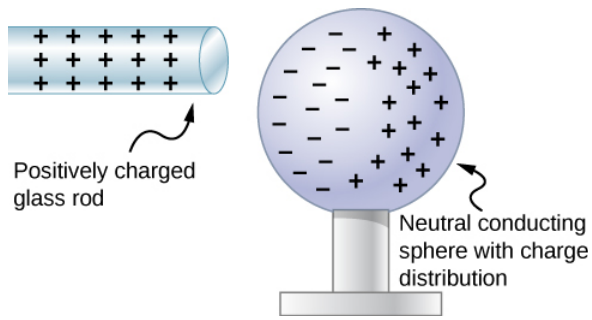


Figure 1.3.2: Induced polarization. A positively charged glass rod is brought near the left side of the conducting sphere, attracting negative charge and leaving the other side of the sphere positively charged. Although the sphere is overall still electrically neutral, it now has a charge distribution, so it can exert an electric force on other nearby charges. Furthermore, the distribution is such that it will be attracted to the glass rod.

The result is the formation of what is called an electric **dipole**, from a Latin phrase meaning "two ends." The presence of electric charges on the insulator—and the electric forces they apply to the conduction electrons—creates, or "induces," the dipole in the conductor.

Neutral objects can be attracted to any charged object. The pieces of straw attracted to polished amber are neutral, for example. If you run a plastic comb through your hair, the charged comb can pick up neutral pieces of paper. Figure 1.3.3 shows how the polarization of atoms and molecules in neutral objects results in their attraction to a charged object.

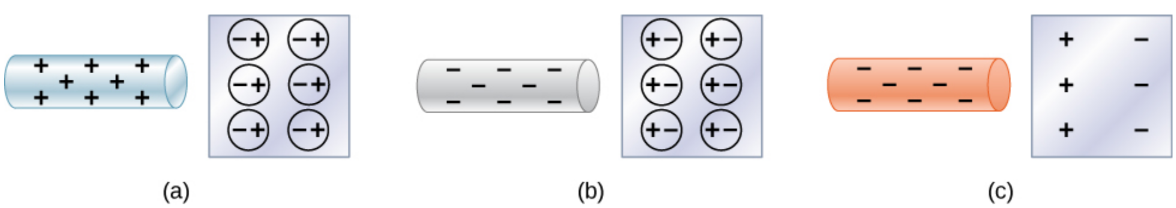


Figure 1.3.3: Both positive and negative objects attract a neutral object by polarizing its molecules. (a) A positive object brought near a neutral insulator polarizes its molecules. There is a slight shift in the distribution of the electrons orbiting the molecule, with unlike charges being brought nearer and like charges moved away. Since the electrostatic force decreases with distance, there is a net attraction. (b) A negative object produces the opposite polarization, but again attracts the neutral object. (c) The same effect occurs for a conductor; since the unlike charges are closer, there is a net attraction.

When a charged rod is brought near a neutral substance, an insulator in this case, the distribution of charge in atoms and molecules is shifted slightly. Opposite charge is attracted nearer the external charged rod, while like charge is repelled. Since the electrostatic force decreases with distance, the repulsion of like charges is weaker than the attraction of unlike charges, and so there is a net attraction. Thus, a positively charged glass rod attracts neutral pieces of paper, as will a negatively charged rubber rod. Some molecules, like water, are polar molecules. Polar molecules have a natural or inherent separation of charge, although they are neutral overall. Polar molecules are particularly affected by other charged objects and show greater polarization effects than molecules with naturally uniform charge distributions.

When the two ends of a dipole can be separated, this method of **charging by induction** may be used to create charged objects without transferring charge. In Figure 1.3.4, we see two neutral metal spheres in contact with one another but insulated from the rest of the world. A positively charged rod is brought near one of them, attracting negative charge to that side, leaving the other sphere positively charged.

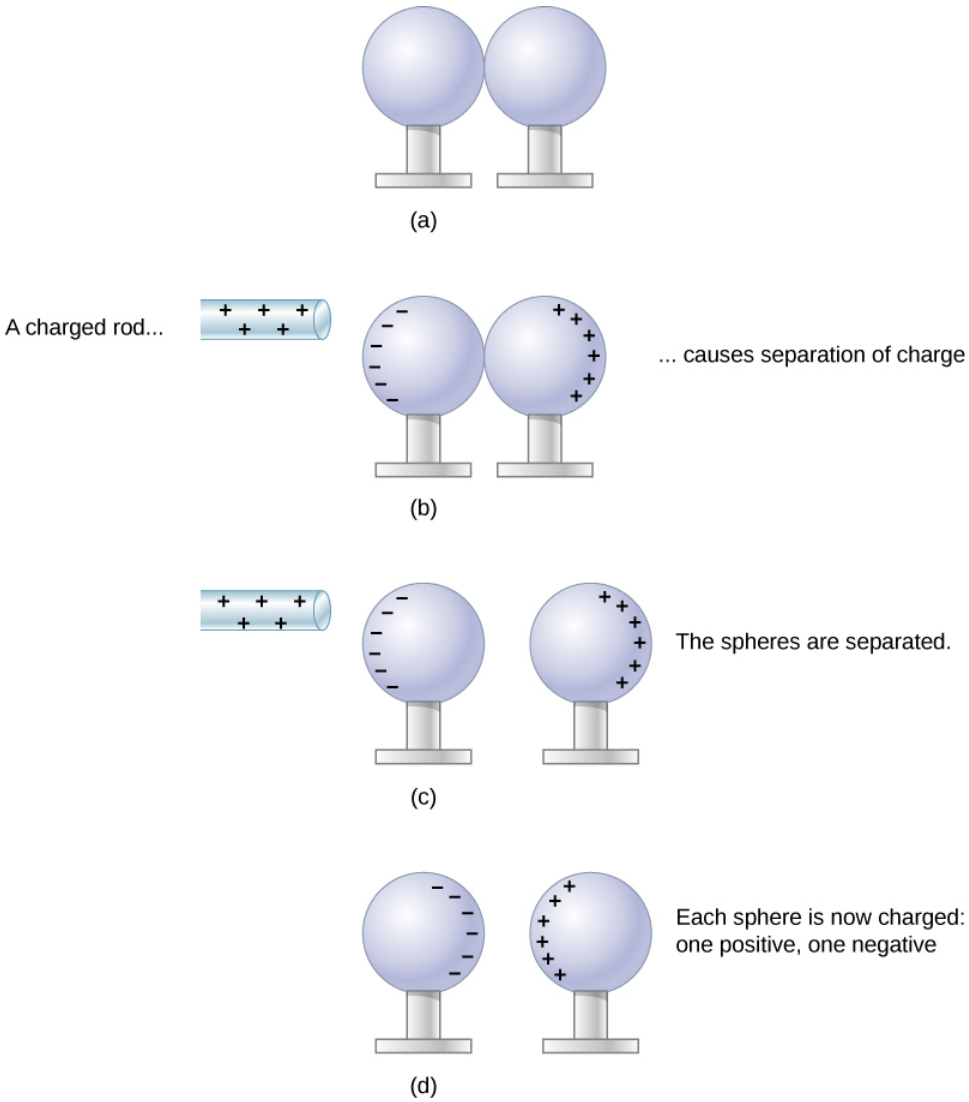


Figure 1.3.4: Charging by induction. (a) Two uncharged or neutral metal spheres are in contact with each other but insulated from the rest of the world. (b) A positively charged glass rod is brought near the sphere on the left, attracting negative charge and leaving the other sphere positively charged. (c) The spheres are separated before the rod is removed, thus separating negative and positive charges. (d) The spheres retain net charges after the inducing rod is removed—without ever having been touched by a charged object.

Another method of charging by induction is shown in Figure 1.3.5. The neutral metal sphere is polarized when a charged rod is brought near it. The sphere is then grounded, meaning that a conducting wire is run from the sphere to the ground. Since Earth is large and most of the ground is a good conductor, it can supply or accept excess charge easily. In this case, electrons are attracted to the sphere through a wire called the ground wire, because it supplies a conducting path to the ground. The ground connection is broken before the charged rod is removed, leaving the sphere with an excess charge opposite to that of the rod. Again, an opposite charge is achieved when charging by induction, and the charged rod loses none of its excess charge.

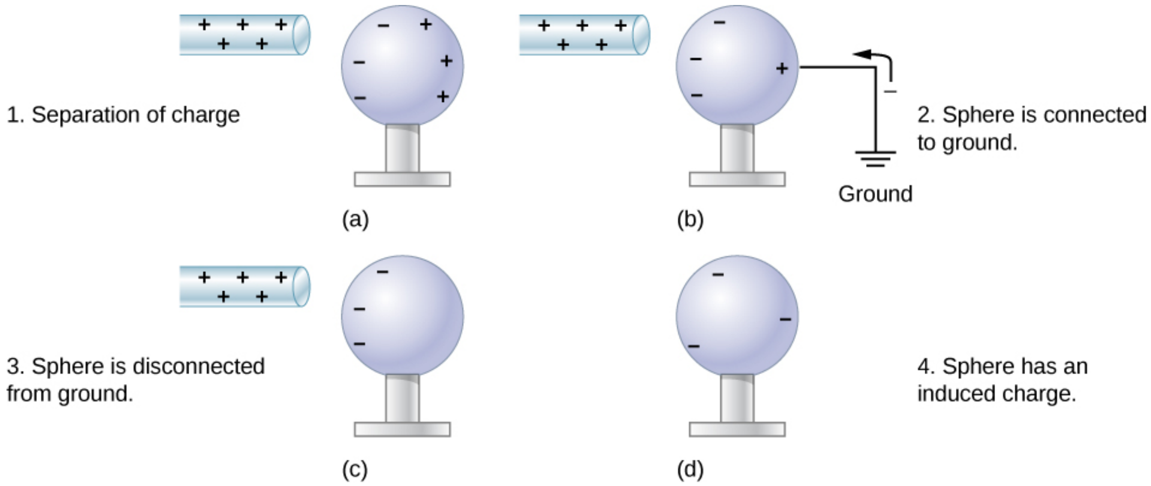


Figure 1.3.5: Charging by induction using a ground connection. (a) A positively charged rod is brought near a neutral metal sphere, polarizing it. (b) The sphere is grounded, allowing electrons to be attracted from Earth's ample supply. (c) The ground connection is broken. (d) The positive rod is removed, leaving the sphere with an induced negative charge.

CHAPTER OVERVIEW

2: Electric Fields

2.1: Coulomb's Law

Learning Objectives

By the end of this section, you will be able to:

- Describe the electric force, both qualitatively and quantitatively
- Calculate the force that charges exert on each other
- Determine the direction of the electric force for different source charges
- Correctly describe and apply the superposition principle for multiple source charges

Experiments with electric charges have shown that if two objects each have electric charge, then they exert an electric force on each other. The magnitude of the force is linearly proportional to the net charge on each object and inversely proportional to the square of the distance between them. (Interestingly, the force does not depend on the mass of the objects.) The direction of the force vector is along the imaginary line joining the two objects and is dictated by the signs of the charges involved.

Let

- $q_1, q_2 =$ the net electric charge of the two objects;
- $\vec{r}_{12} =$ the vector displacement from q_1 to q_2.

The electric force $\vec{F}$ on one of the charges is proportional to the magnitude of its own charge and the magnitude of the other charge, and is inversely proportional to the square of the distance between them:

$$F \propto \frac{q_1 q_2}{r_{12}^2}.$$

This proportionality becomes an equality with the introduction of a proportionality constant. For reasons that will become clear in a later chapter, the proportionality constant that we use is actually a collection of constants. (We discuss this constant shortly.)

Coulomb's Law

The magnitude of the electric force (or **Coulomb force**) between two electrically charged particles is equal to

$$|\mathbf{F}_{12}| = \frac{1}{4\pi\varepsilon_0} \frac{|q_1 q_2|}{r_{12}^2} \tag{2.1.1}$$

The unit vector r has a magnitude of 1 and points along the axis as the charges. If the charges have the same sign, the force is in the same direction as r showing a repelling force. If the charges have different signs, the force is in the opposite direction of r showing an attracting force. (Figure $2.1.1$).

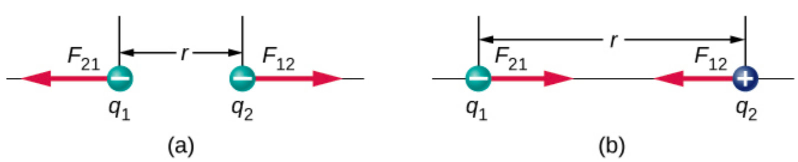


Figure $2.1.1$: The electrostatic force $\vec{F}$ between point charges q_1 and q_2 separated by a distance r is given by Coulomb's law. Note that Newton's third law (every force exerted creates an equal and opposite force) applies as usual—the force on q_1 is equal in magnitude and opposite in direction to the force it exerts on q_2. (a) Like charges; (b) unlike charges.

It is important to note that the electric force is not constant; it is a function of the separation distance between the two charges. If either the test charge or the source charge (or both) move, then $\vec{r}$ changes, and therefore so does the force. An immediate consequence of this is that direct application of Newton's laws with this force can be mathematically difficult, depending on the specific problem at hand. It can (usually) be done, but we almost always look for easier methods of calculating whatever physical quantity we are interested in. (Conservation of energy is the most common choice.)

Finally, the new constant ϵ_0 in Coulomb's law is called the *permittivity of free space*, or (better) the **permittivity of vacuum**. It has a very important physical meaning that we will discuss in a later chapter; for now, it is simply an empirical proportionality constant. Its numerical value (to three significant figures) turns out to be

$$\epsilon_0 = 8.85 \times 10^{-12} \, \frac{\mathrm{C}^2}{\mathrm{N} \cdot \mathrm{m}^2}.$$

These units are required to give the force in Coulomb's law the correct units of newtons. Note that in Coulomb's law, the permittivity of vacuum is only part of the proportionality constant. For convenience, we often define a Coulomb's constant:

$$k_e = \frac{1}{4\pi\epsilon_0} = 8.99 \times 10^9 \, \frac{\mathrm{N} \cdot \mathrm{m}^2}{\mathrm{C}^2}.$$

Example 2.1.1: The Force on the Electron in Hydrogen

A hydrogen atom consists of a single proton and a single electron. The proton has a charge of $+e$ and the electron has $-e$. In the "ground state" of the atom, the electron orbits the proton at most probable distance of $5.29 \times 10^{-11} \mathrm{m}$ (Figure $2.1.2$). Calculate the electric force on the electron due to the proton.

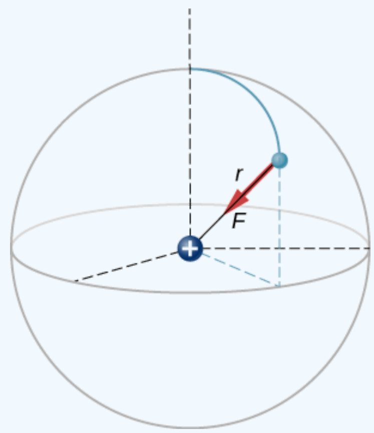


Figure $2.1.2$: A schematic depiction of a hydrogen atom, showing the force on the electron. This depiction is only to enable us to calculate the force; the hydrogen atom does not really look like this.

Strategy

For the purposes of this example, we are treating the electron and proton as two point particles, each with an electric charge, and we are told the distance between them; we are asked to calculate the force on the electron. We thus use Coulomb's law (Equation 2.1.1).

Solution

Our two charges are,

$$\begin{aligned} q_1 &= +e \\ &= +1.602 \times 10^{-19} \, \mathrm{C} \\ q_2 &= -e \\ &= -1.602 \times 10^{-19} \, \mathrm{C} \end{aligned}$$

and the distance between them

$$r = 5.29 \times 10^{-11} \, \mathrm{m}.$$

The magnitude of the force on the electron (Equation 2.1.1) is

$$
\begin{aligned}
F &= \frac{1}{4\pi\epsilon_0}\frac{|q_1q_2|}{r_{12}^2} \\
&= \frac{1}{4\pi\left(8.85\times10^{-12}\,\frac{\mathrm{C}^2}{\mathrm{N}\cdot\mathrm{m}^2}\right)}\frac{(1.602\times10^{-19}\,\mathrm{C})^2}{(5.29\times10^{-11}\,\mathrm{m})^2} \\
&= 8.25\times10^{-8}\,\mathrm{N}.
\end{aligned}
$$

As for the direction, since the charges on the two particles are opposite, the force is attractive; the force on the electron points radially directly toward the proton, everywhere in the electron's orbit. The force is thus expressed as

$$\vec{F} = (8.25\times10^{-8}\,\mathrm{N})\hat{r}.$$

Significance

This is a three-dimensional system, so the electron (and therefore the force on it) can be anywhere in an imaginary spherical shell around the proton. In this "classical" model of the hydrogen atom, the electrostatic force on the electron points in the inward centripetal direction, thus maintaining the electron's orbit. But note that the quantum mechanical model of hydrogen (discussed in Quantum Mechanics) is utterly different.

? Exercise 2.1.1

What would be different if the electron also had a positive charge?

Answer

The force would point outward.

Multiple Source Charges

The analysis that we have done for two particles can be extended to an arbitrary number of particles; we simply repeat the analysis, two charges at a time. Specifically, we ask the question: Given N charges (which we refer to as source charge), what is the net electric force that they exert on some other point charge (which we call the test charge)? Note that we use these terms because we can think of the test charge being used to test the strength of the force provided by the source charges.

Like all forces that we have seen up to now, the net electric force on our test charge is simply the vector sum of each individual electric force exerted on it by each of the individual test charges. Thus, we can calculate the net force on the test charge Q by calculating the force on it from each source charge, taken one at a time, and then adding all those forces together (as vectors). This ability to simply add up individual forces in this way is referred to as the **principle of superposition**, and is one of the more important features of the electric force. In mathematical form, this becomes

$$\vec{F}(r) = \frac{1}{4\pi\epsilon_0}Q\sum_{i=1}^{N}\frac{q_i}{r_i^2}\hat{r}_i^2. \tag{2.1.2}$$

In this expression, Q represents the charge of the particle that is experiencing the electric force $\vec{F}$, and is located at $\vec{r}$ from the origin; the q_i's are the N source charges, and the vectors $\vec{r}_i = r_i\hat{r}_i$ are the displacements from the position of the ith charge to the position of Q. Each of the N unit vectors points directly from its associated source charge toward the test charge. All of this is depicted in Figure $2.1.3$. Please note that there is no physical difference between Q and q_i; the difference in labels is merely to allow clear discussion, with Q being the charge we are determining the force on.

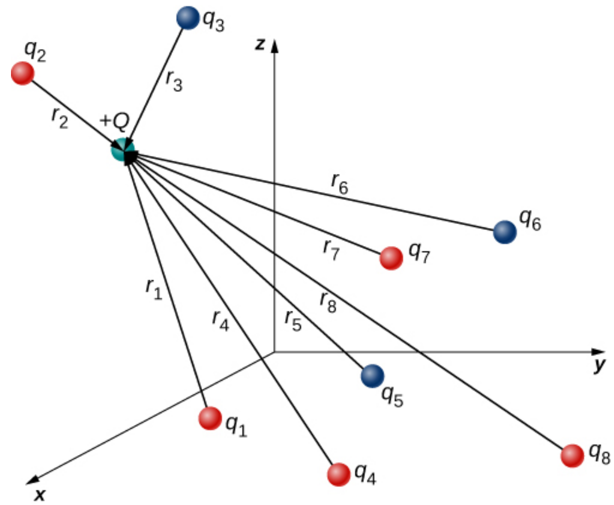


Figure 2.1.3: The eight source charges each apply a force on the single test charge Q. Each force can be calculated independently of the other seven forces. This is the essence of the superposition principle.

(Note that the force vector $\vec{F}_i$ does not necessarily point in the same direction as the unit vector $\hat{r}_i$; it may point in the opposite direction, $-\hat{r}_i$. The signs of the source charge and test charge determine the direction of the force on the test charge.)

There is a complication, however. Just as the source charges each exert a force on the test charge, so too (by Newton's third law) does the test charge exert an equal and opposite force on each of the source charges. As a consequence, each source charge would change position. However, by Equation 2.1.1, the force on the test charge is a function of position; thus, as the positions of the source charges change, the net force on the test charge necessarily changes, which changes the force, which again changes the positions. Thus, the entire mathematical analysis quickly becomes intractable. Later, we will learn techniques for handling this situation, but for now, we make the simplifying assumption that the source charges are fixed in place somehow, so that their positions are constant in time. (The test charge is allowed to move.) With this restriction in place, the analysis of charges is known as **electrostatics**, where "statics" refers to the constant (that is, static) positions of the source charges and the force is referred to as an **electrostatic force**.

✓ Example 2.1.2: The Net Force from Two Source Charges

Three different, small charged objects are placed as shown in Figure 2.1.4. The charges q_1 and q_3 are fixed in place; q_2 is free to move. Given $q_1 = 2e$, $q_2 = -3e$, and $q_3 = -5e$, and that $d = 2.0 \times 10^{-7}\ \text{m}$, what is the net force on the middle charge q_2?

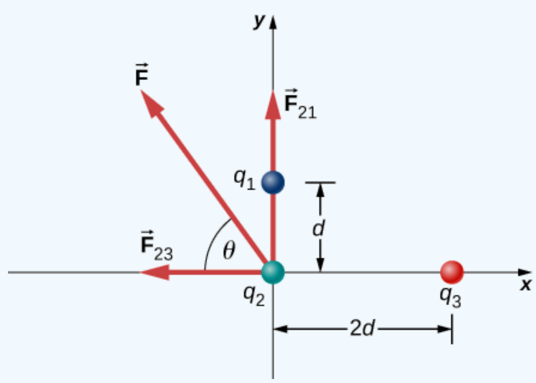


Figure 2.1.4: Source charges q_1 and q_3 each apply a force on q_2.

Strategy

We use Coulomb's law again. The way the question is phrased indicates that q_2 is our test charge, so that q_1 and q_3 are source charges. The principle of superposition says that the force on q_2 from each of the other charges is unaffected by the presence of the other charge. Therefore, we write down the force on q_2 from each and add them together as vectors.

Solution

We have two source charges q_1 and q_3 a test charge q_2, distances r_{21} and r_{23} and we are asked to find a force. This calls for Coulomb's law and superposition of forces. There are two forces:

$$\vec{F} = \vec{F}_{21} + \vec{F}_{23} = \frac{1}{4\pi\epsilon_0}\left[\frac{q_2 q_1}{r_{21}^2}\hat{j} + \left(-\frac{q_2 q_3}{r_{23}^2}\hat{i}\right)\right].$$

We cannot add these forces directly because they don't point in the same direction: $\vec{F}_{12}$ points only in the $-x$-direction, while $\vec{F}_{13}$ points only in the $+y$-direction. The net force is obtained from applying the Pythagorean theorem to its x- and y-components:

$$F = \sqrt{F_x^2 + F_y^2}$$

and

$$\begin{aligned} F_x &= -F_{23} = -\frac{1}{4\pi\epsilon_0}\frac{q_2 q_3}{r_{23}^2} \\ &= -\left(8.99 \times 10^9 \frac{\mathrm{N \cdot m^2}}{\mathrm{C^2}}\right)\frac{(4.806 \times 10^{-19}\,\mathrm{C})(8.01 \times 10^{-19}\,\mathrm{C})}{(4.00 \times 10^{-7}\,\mathrm{m})^2} \\ &= -2.16 \times 10^{-14}\,\mathrm{N} \end{aligned}$$

and

$$\begin{aligned} F_y &= F_{21} = \frac{1}{4\pi\epsilon_0}\frac{q_2 q_1}{r_{21}} \\ &= \left(9.99 \times 10^9 \frac{\mathrm{N \cdot m^2}}{\mathrm{C^2}}\right)\frac{(4.806 \times 10^{-19}\,\mathrm{C})(3.204 \times 10^{-19}\,\mathrm{C})}{(2.00 \times 10^{-7}\,\mathrm{m})^2} \\ &= 3.46 \times 10^{-14}\,\mathrm{N}. \end{aligned}$$

We find that

$$\begin{aligned} F &= \sqrt{F_x^2 + F_y^2} \\ &= 4.08 \times 10^{-14}\,\mathrm{N} \end{aligned}$$

at an angle of

$$\begin{aligned} \phi &= \tan^{-1}\left(\frac{F_y}{F_x}\right) \\ &= \tan^{-1}\left(\frac{3.46 \times 10^{-14}\,\mathrm{N}}{-2.16 \times 10^{-14}\,\mathrm{N}}\right) \\ &= -58^\circ, \end{aligned}$$

that is, 58° above the $-x$-axis, as shown in the diagram.

Significance

Notice that when we substituted the numerical values of the charges, we did not include the negative sign of either q_1 or q_3. Recall that negative signs on vector quantities indicate a reversal of direction of the vector in question. But for electric forces, the direction of the force is determined by the types (signs) of both interacting charges; we determine the force directions by considering whether the signs of the two charges are the same or are opposite. If you also include negative signs from negative charges when you substitute numbers, you run the risk of mathematically reversing the direction of the force you are calculating. Thus, the safest thing to do is to calculate just the magnitude of the force, using the absolute values of the charges, and determine the directions physically.

It's also worth noting that the only new concept in this example is how to calculate the electric forces; everything else (getting the net force from its components, breaking the forces into their components, finding the direction of the net force) is the same as force problems you have done earlier.

? Exercise $2.1.2$

What would be different in Example $2.1.2$ if q_1 were negative rather than positive?

Answer

The net force would point $58°$ below the $-x$-axis.

2.2: Electric Field

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the purpose of the electric field concept
- Describe the properties of the electric field
- Calculate the field of a collection of source charges of either sign

As we showed in the preceding section, the net electric force on a test charge is the vector sum of all the electric forces acting on it, from all of the various source charges, located at their various positions. But what if we use a different test charge, one with a different magnitude, or sign, or both? Or suppose we have a dozen different test charges we wish to try at the same location? We would have to calculate the sum of the forces from scratch. Fortunately, it is possible to define a quantity, called the **electric field**, which is independent of the test charge. It only depends on the configuration of the source charges, and once found, allows us to calculate the force on any test charge.

Defining a Field

Suppose we have N source charges $q_1, q_2, q_3, \ldots q_N$ located at positions $\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots \vec{r}_N$, applying N electrostatic forces on a test charge Q. The net force on Q is

$$\vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \ldots + \vec{F}_N \tag{2.2.1}$$

$$= \frac{1}{4\pi\epsilon_0}\left(\frac{Qq_1}{r_1^2}\hat{r}_1 + \frac{Qq_2}{r_2^2}\hat{r}_2 + \frac{Qq_3}{r_3^2}\hat{r}_3 + \ldots + \frac{Qq_N}{r_1^2}\hat{r}_N\right) \tag{2.2.2}$$

$$= Q\left[\frac{1}{4\pi\epsilon_0}\left(\frac{q_1}{r_1^2}\hat{r}_1 + \frac{q_2}{r_2^2}\hat{r}_2 + \frac{q_3}{r_3^2}\hat{r}_3 + \ldots + \frac{q_N}{r_1^2}\hat{r}_N\right)\right] \tag{2.2.3}$$

We can rewrite this as

$$\vec{F} = Q\vec{E} \tag{2.2.4}$$

where

$$\vec{E} \equiv \frac{1}{4\pi\epsilon_0}\left(\frac{q_1}{r_1^2}\hat{r}_1 + \frac{q_2}{r_2^2}\hat{r}_2 + \frac{q_3}{r_3^2}\hat{r}_3 + \ldots + \frac{q_N}{r_1^2}\hat{r}_N\right) \tag{2.2.5}$$

or, more compactly,

$$\vec{E}(P) \equiv \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{q_i}{r_i^2}\hat{r}_i. \tag{2.2.6}$$

This expression is called the electric field at position $P = P(x, y, z)$ of the N source charges. Here, P is the location of the point in space where you are calculating the field and is relative to the positions $\vec{r}_i$ of the source charges (Figure $2.2.1$). Note that we have to impose a coordinate system to solve actual problems.

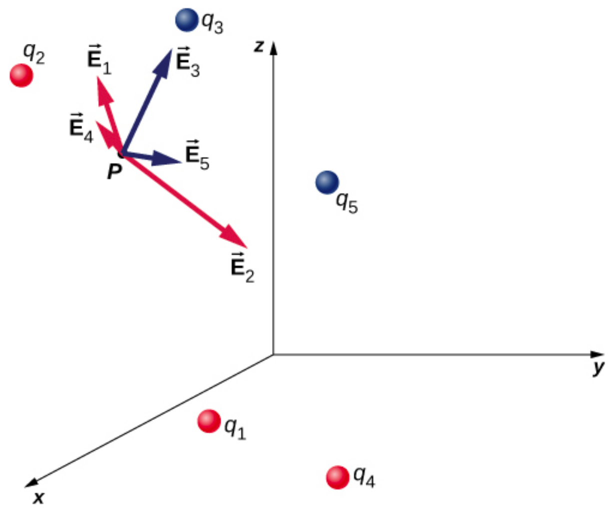


Figure 2.2.1: Each of these five source charges creates its own electric field at every point in space; shown here are the field vectors at an arbitrary point P. Like the electric force, the net electric field obeys the superposition principle.

Notice that the calculation of the electric field makes no reference to the test charge. Thus, the physically useful approach is to calculate the electric field and then use it to calculate the force on some test charge later, if needed. Different test charges experience different forces Equation 2.2.4, but it is the same electric field Equation 2.2.6. That being said, recall that there is no fundamental difference between a test charge and a source charge; these are merely convenient labels for the system of interest. Any charge produces an electric field; however, just as Earth's orbit is not affected by Earth's own gravity, a charge is not subject to a force due to the electric field it generates. Charges are only subject to forces from the electric fields of other charges.

In this respect, the electric field $\vec{E}$ of a point charge is similar to the gravitational field $\vec{g}$ of Earth; once we have calculated the gravitational field at some point in space, we can use it any time we want to calculate the resulting force on any mass we choose to place at that point. In fact, this is exactly what we do when we say the gravitational field of Earth (near Earth's surface) has a value of $9.81\ \mathrm{m/s^2}$ and then we calculate the resulting force (i.e., weight) on different masses. Also, the general expression for calculating $\vec{g}$ at arbitrary distances from the center of Earth (i.e., not just near Earth's surface) is very similar to the expression for E:

$$\vec{g} = G\frac{M}{r^2}\hat{r}$$

where G is a proportionality constant, playing the same role for $\vec{g}$ as $\dfrac{1}{4\pi\epsilon_0}$ does for $\vec{E}$. The value of $\vec{g}$ is calculated once and is then used in an endless number of problems.

To push the analogy further, notice the units of the electric field: From $F = QE$, the units of **E** are newtons per coulomb, N/C, that is, the electric field applies a force on each unit charge. Now notice the units of g: From $w = mg$ the units of g are newtons per kilogram, N/kg, that is, the gravitational field applies a force on each unit mass. We could say that the gravitational field of Earth, near Earth's surface, has a value of 9.81 N/kg.

The Meaning of "Field"

Recall from your studies of gravity that the word "field" in this context has a precise meaning. A field, in physics, is a physical quantity whose value depends on (is a function of) position, relative to the source of the field. In the case of the electric field, Equation 2.2.6 shows that the value of $\vec{E}$ (both the magnitude and the direction) depends on where in space the point P is located, measured from the locations $\vec{r}_i$ of the source charges q_i.

In addition, since the electric field is a vector quantity, the electric field is referred to as a *vector field*. (The gravitational field is also a vector field.) In contrast, a field that has only a magnitude at every point is a *scalar field*. The temperature in a room is an example of a scalar field. It is a field because the temperature, in general, is different at different locations in the room, and it is a scalar field because temperature is a scalar quantity.

Also, as you did with the gravitational field of an object with mass, you should picture the electric field of a charge-bearing object (the source charge) as a continuous, immaterial substance that surrounds the source charge, filling all of space—in principle, to $\pm\infty$ in all directions. The field exists at every physical point in space. To put it another way, the electric charge on an object alters the space around the charged object in such a way that all other electrically charged objects in space experience an electric force as a result of being in that field. The electric field, then, is the mechanism by which the electric properties of the source charge are transmitted to and through the rest of the universe. (Again, the range of the electric force is infinite.)

We will see in subsequent chapters that the speed at which electrical phenomena travel is the same as the speed of light. There is a deep connection between the electric field and light.

Superposition

Yet another experimental fact about the field is that it obeys the superposition principle. In this context, that means that we can (in principle) calculate the total electric field of many source charges by calculating the electric field of only q_1 at position **P**, then calculate the field of q_2 at **P**, while—and this is the crucial idea—ignoring the field of, and indeed even the existence of, q_1. We can repeat this process, calculating the field of each individual source charge, independently of the existence of any of the other charges. The total electric field, then, is the vector sum of all these fields. That, in essence, is what Equation 2.2.6 says.

In the next section, we describe how to determine the shape of an electric field of a source charge distribution and how to sketch it.

The Direction of the Field

Equation 2.2.6 enables us to determine the magnitude of the electric field, but we need the direction also. We use the convention that the direction of any electric field vector is the same as the direction of the electric force vector that the field would apply to a positive test charge placed in that field. Such a charge would be repelled by positive source charges (the force on it would point away from the positive source charge) but attracted to negative charges (the force points toward the negative source).

Direction of the Electric Field

By convention, all electric fields $\vec{E}$ point away from positive source charges and point toward negative source charges.

Phet Simulation: Electric Field of Dreams

Download the Electric Field of Dreams PhET simulation and add charges to the and see how they react to the electric field. Turn on a background electric field and adjust the direction and magnitude.

Example $2.2.1A$: The E-field of an Atom

In an ionized helium atom, the most probable distance between the nucleus and the electron is $r = 26.5 \times 10^{-12}\,\mathrm{m}$. What is the electric field due to the nucleus at the location of the electron?

Strategy

Note that although the electron is mentioned, it is not used in any calculation. The problem asks for an electric field, not a force; hence, there is only one charge involved, and the problem specifically asks for the field due to the nucleus. Thus, the electron is a red herring; only its distance matters. Also, since the distance between the two protons in the nucleus is much,

much smaller than the distance of the electron from the nucleus, we can treat the two protons as a single charge +2e (Figure 2.2.2).

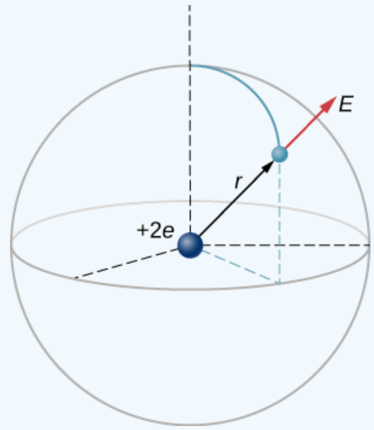


Figure 2.2.2: A schematic representation of a helium atom. Again, helium physically looks nothing like this, but this sort of diagram is helpful for calculating the electric field of the nucleus.

Solution

The electric field is calculated by

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{q_i}{r_i^2}\hat{r}_i.$$

Since there is only one source charge (the nucleus), this expression simplifies to

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}.$$

Here, $q = 2e = 2(1.6 \times 10^{-19}\ \text{C})$ (since there are two protons) and r is given; substituting gives

$$\begin{aligned}\vec{E} &= \frac{1}{4\pi\left(8.85 \times 10^{-12}\frac{\text{C}^2}{\text{N}\cdot\text{m}^2}\right)}\frac{2(1.6 \times 10^{-19}\ \text{C})}{26.5 \times 10^{-12}\ \text{m})^2}\hat{r} \\ &= 4.1 \times 10^{12}\frac{\text{N}}{\text{C}}\hat{r}.\end{aligned}$$

The direction of $\vec{E}$ is radially away from the nucleus in all directions. Why? Because a positive test charge placed in this field would accelerate radially away from the nucleus (since it is also positively charged), and again, the convention is that the direction of the electric field vector is defined in terms of the direction of the force it would apply to positive test charges.

Example 2.2.1B: The E-Field above Two Equal Charges

a. Find the electric field (magnitude and direction) a distance z above the midpoint between two equal charges $+q$ that are a distance d apart (Figure 2.2.3). Check that your result is consistent with what you'd expect when $z \gg d$.
b. The same as part (a), only this time make the right-hand charge $-q$ instead of $+q$.

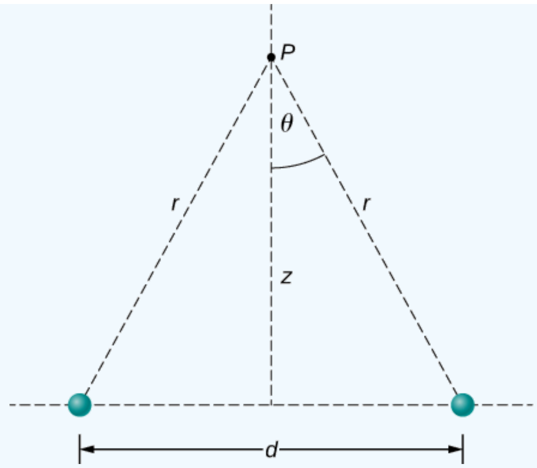


Figure 2.2.3: Finding the field of two identical source charges at the point P. Due to the symmetry, the net field at P is entirely vertical. (Notice that this is **not** true away from the midline between the charges.)

Strategy

We add the two fields as vectors, per Equation 2.2.6. Notice that the system (and therefore the field) is symmetrical about the vertical axis; as a result, the horizontal components of the field vectors cancel. This simplifies the math. Also, we take care to express our final answer in terms of only quantities that are given in the original statement of the problem: q, z, d, and constants (π, ϵ_0).

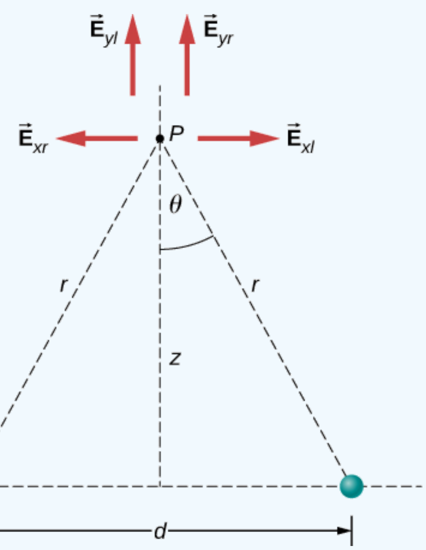


Figure 2.2.4. Note that the horizontal components of the electric fields from the two charges cancel each other out, while the vertical components add together.

Solution

a. By symmetry, the horizontal (x)-components of $\vec{E}$ cancel (Figure 2.2.4);

$$\begin{aligned} E_x &= \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\sin\theta - \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\sin\theta \\ &= 0. \end{aligned}$$

The vertical (z)-component is given by

$$\begin{aligned} E_z &= \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\cos\theta + \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\cos\theta \\ &= \frac{1}{4\pi\epsilon_0}\frac{2q}{r^2}\cos\theta. \end{aligned}$$

Since none of the other components survive, this is the entire electric field, and it points in the $\hat{k}$ direction. Notice that this calculation uses the principle of **superposition**; we calculate the fields of the two charges independently and then add them together.

What we want to do now is replace the quantities in this expression that we don't know (such as r), or can't easily measure (such as $\cos\theta$ with quantities that we do know, or can measure. In this case, by geometry,

$$r^2 = z^2 + \left(\frac{d}{2}\right)^2$$

and

$$\cos\theta = \frac{z}{R} = \frac{z}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{1/2}}.$$

Thus, substituting,

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{2q}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^2}\frac{z}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{1/2}}\hat{k}.$$

Simplifying, the desired answer is

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{2qz}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{3/2}}\hat{k}. \tag{2.2.7}$$

b. If the source charges are equal and opposite, the vertical components cancel because

$$E_z = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\,cos\,\theta - \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\cos\theta = 0$$

and we get, for the horizontal component of $\vec{E}$.

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{qd}{\left[z^2 + \left(\frac{d}{2}\right)^2\right]^{3/2}}\hat{i}. \tag{2.2.8}$$

Significance

It is a very common and very useful technique in physics to check whether your answer is reasonable by evaluating it at extreme cases. In this example, we should evaluate the field expressions for the cases $d = 0$, $z \gg d$, and $z \to \infty$, and confirm that the resulting expressions match our physical expectations. Let's do so:

Let's start with Equation 2.2.7, the field of two identical charges. From far away (i.e., $z >> d$), the two source charges should "merge" and we should then "see" the field of just one charge, of size $2q$. So, let $z \gg d$; then we can neglect d^2 in Equation 2.2.7 to obtain

$$\begin{aligned}\lim_{d\to 0}\vec{E} &= \frac{1}{4\pi\epsilon_0}\frac{2qz}{[z^2]^{3/2}}\hat{k}\\ &= \frac{1}{4\pi\epsilon_0}\frac{2qz}{z^3}\hat{k}\\ &= \frac{1}{4\pi\epsilon_0}\frac{2q}{z^2}\hat{k},\end{aligned}$$

which is the correct expression for a field at a distance z away from a charge $2q$.

Next, we consider the field of equal and opposite charges, Equation 2.2.8. It can be shown (via a Taylor expansion) that for $d \ll z \ll \infty$, , this becomes

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{qd}{z^3}\hat{i},$$

which is the field of a dipole, a system that we will study in more detail later. (Note that the units of $\vec{E}$ are still correct in this expression, since the units of d in the numerator cancel the unit of the "extra" z in the denominator.) If z is *very* large $(z \to \infty)$, then $E \to 0$, as it should; the two charges "merge" and so cancel out.

Exercise 2.2.1

What is the electric field due to a single point particle?

Answer

$$\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$$

Phe Simulation: Electric Field Hockey

Try this simulation of electric field hockey to get the charge in the goal by placing other charges on the field.

2.3: Calculating Electric Fields of Charge Distributions

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain what a continuous source charge distribution is and how it is related to the concept of quantization of charge
- Describe line charges, surface charges, and volume charges
- Calculate the field of a continuous source charge distribution of either sign

The charge distributions we have seen so far have been discrete: made up of individual point particles. This is in contrast with a **continuous charge distribution**, which has at least one nonzero dimension. If a charge distribution is continuous rather than discrete, we can generalize the definition of the electric field. We simply divide the charge into infinitesimal pieces and treat each piece as a point charge.

Note that because charge is quantized, there is no such thing as a "truly" continuous charge distribution. However, in most practical cases, the total charge creating the field involves such a huge number of discrete charges that we can safely ignore the discrete nature of the charge and consider it to be continuous. This is exactly the kind of approximation we make when we deal with a bucket of water as a continuous fluid, rather than a collection of H_2O molecules.

Our first step is to define a charge density for a charge distribution along a line, across a surface, or within a volume, as shown in Figure $2.3.1$.

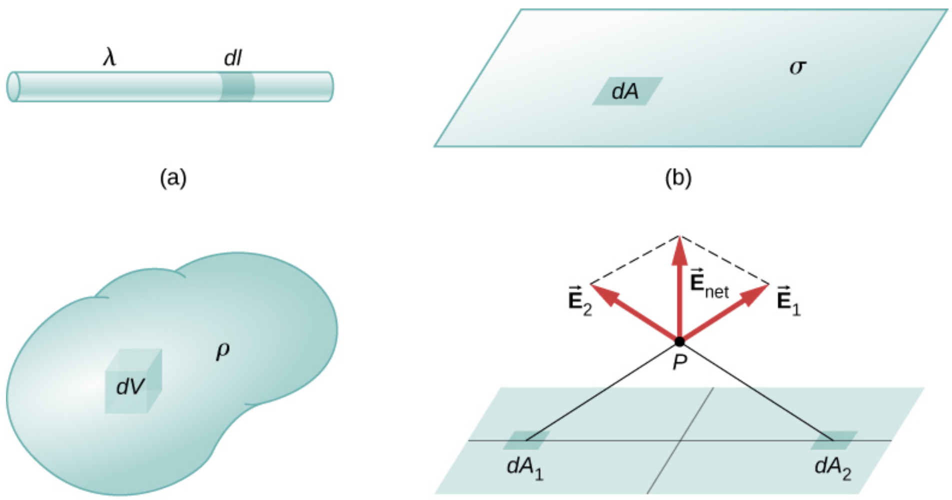


Figure $2.3.1$: The configuration of charge differential elements for a (a) line charge, (b) sheet of charge, and (c) a volume of charge. Also note that (d) some of the components of the total electric field cancel out, with the remainder resulting in a net electric field.

Definitions: Charge Densities

Definitions of charge density:

- **linear charge density:** $\lambda \equiv$ charge per unit length (Figure $2.3.1a$); units are coulombs per meter $(\mathrm{C/m})$
- **surface charge density:** $\sigma \equiv$ charge per unit area (Figure $2.3.1b$); units are coulombs per square meter $(\mathrm{C/m^2})$
- **volume charge density:** $\rho \equiv$ charge per unit volume (Figure $2.3.1c$); units are coulombs per square meter $(\mathrm{C/m^3})$

For a line charge, a surface charge, and a volume charge, the summation in the definition of an Electric field discussed previously becomes an integral and q_i is replaced by $dq = \lambda\, dl$, $\sigma\, dA$, or $\rho\, dV$, respectively:

$$\underbrace{\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\left(\frac{q_i}{r^2}\right)\hat{r}}_{\text{Point charges}} \tag{2.3.1}$$

$$\underbrace{\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{line}}\left(\frac{\lambda\,dl}{r^2}\right)\hat{r}}_{\text{Line charge}} \tag{2.3.2}$$

$$\underbrace{\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{surface}}\left(\frac{\sigma\,dA}{r^2}\right)\hat{r}}_{\text{Surface charge}} \tag{2.3.3}$$

$$\underbrace{\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{volume}}\left(\frac{\rho\,dV}{r^2}\right)\hat{r}}_{\text{Volume charge}} \tag{2.3.4}$$

The integrals in Equations 2.3.1-2.3.4 are generalizations of the expression for the field of a point charge. They implicitly include and assume the principle of superposition. The "trick" to using them is almost always in coming up with correct expressions for dl, dA, or dV, as the case may be, expressed in terms of **r**, and also expressing the charge density function appropriately. It may be constant; it might be dependent on location.

Note carefully the meaning of r in these equations: It is the distance from the charge element (q_i, $\lambda\,dl$, $\sigma\,dA$, $\rho\,dV$) to the location of interest, $P(x, y, z)$ (the point in space where you want to determine the field). However, don't confuse this with the meaning of $\hat{r}$; we are using it and the vector notation $\vec{E}$ to write three integrals at once. That is, Equation 2.3.2 is actually

$$E_x(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{line}}\left(\frac{\lambda\,dl}{r^2}\right)_x, \tag{2.3.5}$$

$$E_y(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{line}}\left(\frac{\lambda\,dl}{r^2}\right)_y, \tag{2.3.6}$$

$$E_z(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{line}}\left(\frac{\lambda\,dl}{r^2}\right)_z \tag{2.3.7}$$

Example 2.3.1: Electric Field of a Line Segment

Find the electric field a distance z above the midpoint of a straight line segment of length L that carries a uniform line charge density λ.

Strategy

Since this is a continuous charge distribution, we conceptually break the wire segment into differential pieces of length dl, each of which carries a differential amount of charge

$$dq = \lambda\,dl.$$

Then, we calculate the differential field created by two symmetrically placed pieces of the wire, using the symmetry of the setup to simplify the calculation (Figure 2.3.2). Finally, we integrate this differential field expression over the length of the wire (half of it, actually, as we explain below) to obtain the complete electric field expression.

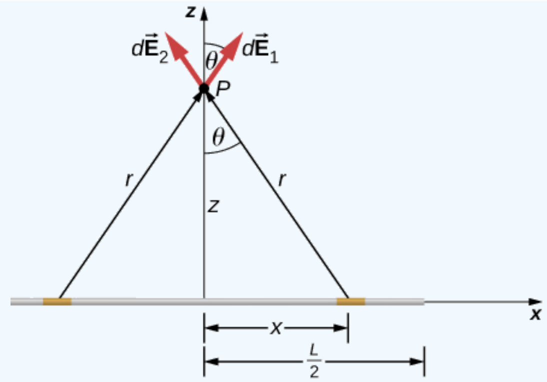


Figure 2.3.2: A uniformly charged segment of wire. The electric field at point P can be found by applying the superposition principle to symmetrically placed charge elements and integrating.

Solution

Before we jump into it, what do we expect the field to “look like” from far away? Since it is a finite line segment, from far away, it should look like a point charge. We will check the expression we get to see if it meets this expectation.

The electric field for a line charge is given by the general expression

$$\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\int_{\text{line}} \frac{\lambda dl}{r^2}\hat{r}.$$

The symmetry of the situation (our choice of the two identical differential pieces of charge) implies the horizontal (x)-components of the field cancel, so that the net field points in the z-direction. Let’s check this formally.

The total field $\vec{E}(P)$ is the vector sum of the fields from each of the two charge elements (call them $\vec{E}_1$ and $\vec{E}_2$, for now):

$$\begin{aligned}\vec{E}(P) &= \vec{E}_1 + \vec{E}_2 \\ &= E_{1x}\hat{i} + E_{1z}\hat{k} + E_{2x}(-\hat{i}) + E_{2z}\hat{k}.\end{aligned}$$

Because the two charge elements are identical and are the same distance away from the point P where we want to calculate the field, $E_{1x} = E_{2x}$, so those components cancel. This leaves

$$\begin{aligned}\vec{E}(P) &= E_{1z}\hat{k} + E_{2z}\hat{k} \\ &= E_1\,\cos\theta\,\hat{k} + E_2\,\cos\theta\,\hat{k}.\end{aligned}$$

These components are also equal, so we have

$$\begin{aligned}\vec{E}(P) &= \frac{1}{4\pi\epsilon_0}\int\frac{\lambda dl}{r^2}\cos\theta\,\hat{k} + \frac{1}{4\pi\epsilon_0}\int\frac{\lambda dl}{r^2}\cos\theta\,\hat{k} \\ &= \frac{1}{4\pi\epsilon_0}\int_0^{L/2}\frac{2\lambda dx}{r^2}\cos\theta\,\hat{k}\end{aligned}$$

where our differential line element dl is dx, in this example, since we are integrating along a line of charge that lies on the x-axis. (The limits of integration are 0 to $\frac{L}{2}$, not $-\frac{L}{2}$ to $+\frac{L}{2}$, because we have constructed the net field from two differential pieces of charge dq. If we integrated along the entire length, we would pick up an erroneous factor of 2.)

In principle, this is complete. However, to actually calculate this integral, we need to eliminate all the variables that are not given. In this case, both r and θ change as we integrate outward to the end of the line charge, so those are the variables to get rid of. We can do that the same way we did for the two point charges: by noticing that

$$r = (z^2 + x^2)^{1/2}$$

and

$$\cos\theta = \frac{z}{r} = \frac{z}{(z^2+x^2)^{1/2}}.$$

Substituting, we obtain

$$\begin{aligned}\vec{E}(P) &= \frac{1}{4\pi\epsilon_0}\int_0^{L/2}\frac{2\lambda dx}{(z^2+x^2)}\frac{z}{(z^2+x^2)^{1/2}}\hat{k} \\ &= \frac{1}{4\pi\epsilon_0}\int_0^{L/2}\frac{2\lambda z}{(z^2+x^2)^{3/2}}dx\,\hat{k} \\ &= \frac{2\lambda z}{4\pi\epsilon_0}\left[\frac{x}{z^2\sqrt{z^2+x^2}}\right]_0^{L/2}\hat{k}.\end{aligned}$$

which simplifies to

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{\lambda L}{z\sqrt{z^2+\frac{L^2}{4}}}\hat{k}. \tag{2.3.8}$$

Significance

Notice, once again, the use of symmetry to simplify the problem. This is a very common strategy for calculating electric fields. The fields of nonsymmetrical charge distributions have to be handled with multiple integrals and may need to be calculated numerically by a computer.

Exercise 2.3.1

How would the strategy used above change to calculate the electric field at a point a distance z above one end of the finite line segment?

Answer

We will no longer be able to take advantage of symmetry. Instead, we will need to calculate each of the two components of the electric field with their own integral.

Example 2.3.2: Electric Field of an Infinite Line of Charge

Find the electric field a distance z above the midpoint of an infinite line of charge that carries a uniform line charge density λ.

Strategy

This is exactly like the preceding example, except the limits of integration will be $-\infty$ to $+\infty$.

Solution

Again, the horizontal components cancel out, so we wind up with

$$\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\int_{-\infty}^{\infty}\frac{\lambda dx}{r^2}\cos\theta\hat{k}$$

where our differential line element dl is dx, in this example, since we are integrating along a line of charge that lies on the x-axis. Again,

$$\begin{aligned}\cos\theta &= \frac{z}{r} \\ &= \frac{z}{(z^2+x^2)^{1/2}}.\end{aligned}$$

Substituting, we obtain

$$\begin{aligned}\vec{E}(P) &= \frac{1}{4\pi\epsilon_0}\int_{-\infty}^{\infty}\frac{\lambda dx}{(z^2+x^2)}\frac{z}{(z^2+x^2)^{1/2}}\hat{k}\\ &= \frac{1}{4\pi\epsilon_0}\int_{-\infty}^{\infty}\frac{\lambda z}{(z^2+x^2)^{3/2}}dx\,\hat{k}\\ &= \frac{1}{4\pi\epsilon_0}\left[\frac{x}{z^2\sqrt{z^2+x^2}}\right]_{-\infty}^{\infty}\hat{k}\end{aligned}$$

which simplifies to

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{2\lambda}{z}\hat{k}.$$

Significance

Our strategy for working with continuous charge distributions also gives useful results for charges with infinite dimension.

In the case of a finite line of charge, note that for $z \gg L$, z^2 dominates the L in the denominator, so that Equation 2.3.8 simplifies to

$$\vec{E} \approx \frac{1}{4\pi\epsilon_0}\frac{\lambda L}{z^2}\hat{k}. \tag{2.3.9}$$

If you recall that $\lambda L = q$ the total charge on the wire, we have retrieved the expression for the field of a point charge, as expected.

In the limit $L \to \infty$ on the other hand, we get the field of an **infinite straight wire**, which is a straight wire whose length is much, much greater than either of its other dimensions, and also much, much greater than the distance at which the field is to be calculated:

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\frac{2\lambda}{z}\hat{k}. \tag{2.3.10}$$

An interesting artifact of this infinite limit is that we have lost the usual $1/r^2$ dependence that we are used to. This will become even more intriguing in the case of an infinite plane.

Example $2.3.3A$: Electric Field due to a Ring of Charge

A ring has a uniform charge density λ, with units of coulomb per unit meter of arc. Find the electric field at a point on the axis passing through the center of the ring.

Strategy

We use the same procedure as for the charged wire. The difference here is that the charge is distributed on a circle. We divide the circle into infinitesimal elements shaped as arcs on the circle and use polar coordinates shown in Figure $2.3.3$.

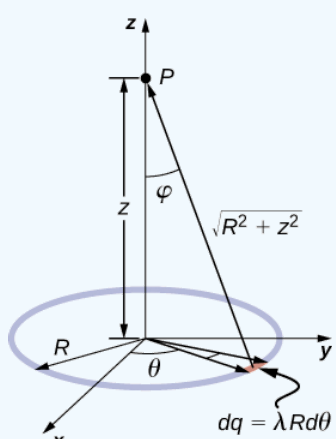


Figure $2.3.3$: The system and variable for calculating the electric field due to a ring of charge.

Solution

The electric field for a line charge is given by the general expression

$$\vec{E}(P)=\frac{1}{4\pi\epsilon_0}\int_{\text{line}}\frac{\lambda dl}{r^2}\hat{r}.$$

A general element of the arc between θ and $\theta+d\theta$ is of length $R\,d\theta$ and therefore contains a charge equal to $\lambda R\,d\theta$. The element is at a distance of $r=\sqrt{z^2+R^2}$ from P, the angle is $\cos\phi=\frac{z}{\sqrt{z^2+R^2}}$ and therefore the electric field is

$$\begin{aligned}\vec{E}(P)&=\frac{1}{4\pi\epsilon_0}\int_{\text{line}}\frac{\lambda dl}{r^2}\hat{r}=\frac{1}{4\pi\epsilon_0}\int_0^{2\pi}\frac{\lambda R d\theta}{z^2+R^2}\frac{z}{\sqrt{z^2+R^2}}\hat{z}\\&=\frac{1}{4\pi\epsilon_0}\frac{\lambda R z}{(z^2+R^2)^{3/2}}\hat{z}\int_0^{2\pi}d\theta\\&=\frac{1}{4\pi\epsilon_0}\frac{2\pi\lambda R z}{(z^2+R^2)^{3/2}}\hat{z}\\&=\frac{1}{4\pi\epsilon_0}\frac{q_{tot}z}{(z^2+R^2)^{3/2}}\hat{z}.\end{aligned}$$

Significance

As usual, symmetry simplified this problem, in this particular case resulting in a trivial integral. Also, when we take the limit of $z\gg R$, we find that

$$\vec{E}\approx\frac{1}{4\pi\epsilon_0}\frac{q_{tot}}{z^2}\hat{z},$$

as we expect.

✓ Example 2.3.3B: The Field of a Disk

Find the electric field of a circular thin disk of radius R and uniform charge density at a distance z above the center of the disk (Figure 2.3.4)

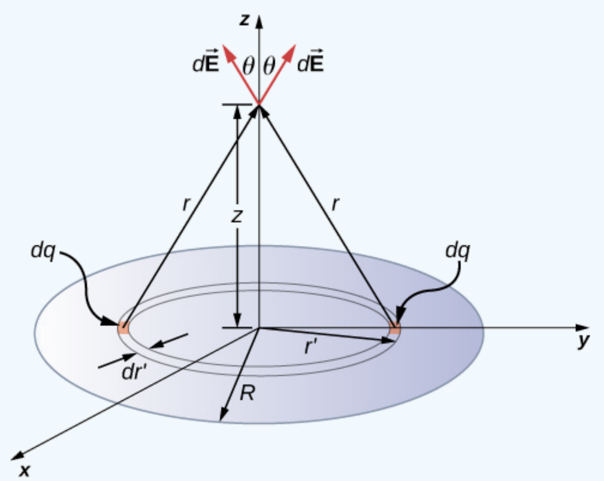


Figure 2.3.4: A uniformly charged disk. As in the line charge example, the field above the center of this disk can be calculated by taking advantage of the symmetry of the charge distribution.

Strategy

The electric field for a surface charge is given by

$$\vec{E}(P)=\frac{1}{4\pi\epsilon_0}\int_{\text{surface}}\frac{\sigma\,dA}{r^2}\hat{r}.$$

To solve surface charge problems, we break the surface into symmetrical differential "stripes" that match the shape of the surface; here, we'll use rings, as shown in the figure. Again, by symmetry, the horizontal components cancel and the field is entirely in the vertical $(\hat{k})$ direction. The vertical component of the electric field is extracted by multiplying by θ, so

$$\vec{E}(P) = \frac{1}{4\pi\epsilon_0}\int_{surface}\frac{\sigma\, dA}{r^2}\cos\theta\,\hat{k}.$$

As before, we need to rewrite the unknown factors in the integrand in terms of the given quantities. In this case,

$$dA = 2\pi r'\,dr' \tag{2.3.11}$$

$$r^2 = r'^2 + z^2 \tag{2.3.12}$$

$$\cos\theta = \frac{z}{(r'^2+z^2)^{1/2}}. \tag{2.3.13}$$

(Please take note of the two different "r's" here; r is the distance from the differential ring of charge to the point P where we wish to determine the field, whereas r' is the distance from the center of the disk to the differential ring of charge.) Also, we already performed the polar angle integral in writing down dA.

Solution

Substituting all this in, we get

$$\begin{aligned}\vec{E}(P) &= \vec{E}(z)\\ &= \frac{1}{4\pi\epsilon_0}\int_0^R\frac{\sigma(2\pi r'\,dr')z}{(r'^2+z^2)^{3/2}}\hat{k}\\ &= \frac{1}{4\pi\epsilon_0}(2\pi\sigma z)\left(\frac{1}{z}-\frac{1}{\sqrt{R^2+z^2}}\right)\hat{k}\end{aligned}$$

or, more simply,

$$\vec{E}(z) = \frac{1}{4\pi\epsilon_0}\left(2\pi\sigma-\frac{2\pi\sigma z}{\sqrt{R^2+z^2}}\right)\hat{k}. \tag{2.3.14}$$

Significance

Again, it can be shown (via a Taylor expansion) that when $z \gg R$, this reduces to

$$\vec{E}(z) \approx \frac{1}{4\pi\epsilon_0}\frac{\sigma\pi R^2}{z^2}\hat{k},$$

which is the expression for a point charge $Q = \sigma\pi R^2$.

? Exercise 2.3.3

How would the above limit change with a uniformly charged rectangle instead of a disk?

Answer

The point charge would be $Q = \sigma ab$ where a and b are the sides of the rectangle but otherwise identical.

As $R \to \infty$, Equation 2.3.14 reduces to the field of an infinite plane, which is a flat sheet whose area is much, much greater than its thickness, and also much, much greater than the distance at which the field is to be calculated:

$$\vec{E} = \lim_{R\to\infty}\frac{1}{4\pi\epsilon_0}\left(2\pi\sigma-\frac{2\pi\sigma z}{\sqrt{R^2+z^2}}\right)\hat{k} \tag{2.3.15}$$

$$= \frac{\sigma}{2\epsilon_0}\hat{k}. \tag{2.3.16}$$

Note that this field is constant. This surprising result is, again, an artifact of our limit, although one that we will make use of repeatedly in the future. To understand why this happens, imagine being placed above an infinite plane of constant charge. Does the plane look any different if you vary your altitude? No—you still see the plane going off to infinity, no matter how far you are from it. It is important to note that Equation $2.3.16$ is because we are above the plane. If we were below, the field would point in the $-\hat{k}$ direction.

Example 2.3.4: The Field of Two Infinite Planes

Find the electric field everywhere resulting from two infinite planes with equal but opposite charge densities (Figure $2.3.5$).

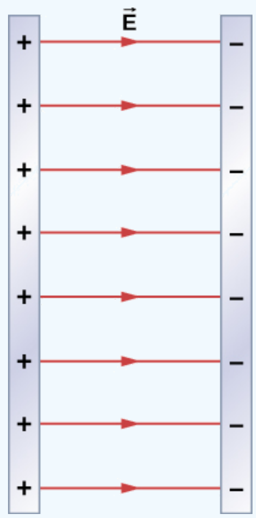


Figure $2.3.5$: Two charged infinite planes. Note the direction of the electric field.

Strategy

We already know the electric field resulting from a single infinite plane, so we may use the principle of superposition to find the field from two.

Solution

The electric field points away from the positively charged plane and toward the negatively charged plane. Since the σ are equal and opposite, this means that in the region outside of the two planes, the electric fields cancel each other out to zero. However, in the region between the planes, the electric fields add, and we get

$$\vec{E} = \frac{\sigma}{\epsilon_0}\hat{i}$$

for the electric field. The $\hat{i}$ is because in the figure, the field is pointing in the $+x$-direction.

Significance

Systems that may be approximated as two infinite planes of this sort provide a useful means of creating uniform electric fields.

Exercise 2.3.1

What would the electric field look like in a system with two parallel positively charged planes with equal charge densities?

Answer

The electric field would be zero in between, and have magnitude $\frac{\sigma}{\epsilon_0}$ everywhere else.

- **5.6: Calculating Electric Fields of Charge Distributions** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/university-physics-volume-2.

2.4: Electric Field Lines

Learning Objectives

By the end of this section, you will be able to:

- Explain the purpose of an electric field diagram
- Describe the relationship between a vector diagram and a field line diagram
- Explain the rules for creating a field diagram and why these rules make physical sense
- Sketch the field of an arbitrary source charge

Now that we have some experience calculating electric fields, let's try to gain some insight into the geometry of electric fields. As mentioned earlier, our model is that the charge on an object (the source charge) alters space in the region around it in such a way that when another charged object (the test charge) is placed in that region of space, that test charge experiences an electric force. The concept of electric **field lines**, and of electric field line diagrams, enables us to visualize the way in which the space is altered, allowing us to visualize the field. The purpose of this section is to enable you to create sketches of this geometry, so we will list the specific steps and rules involved in creating an accurate and useful sketch of an electric field.

It is important to remember that electric fields are three-dimensional. Although in this book we include some pseudo-three-dimensional images, several of the diagrams that you'll see (both here, and in subsequent chapters) will be two-dimensional projections, or cross-sections. Always keep in mind that in fact, you're looking at a three-dimensional phenomenon.

Our starting point is the physical fact that the electric field of the source charge causes a test charge in that field to experience a force. By definition, electric field vectors point in the same direction as the electric force that a (hypothetical) positive test charge would experience, if placed in the field (Figure $2.4.1$).

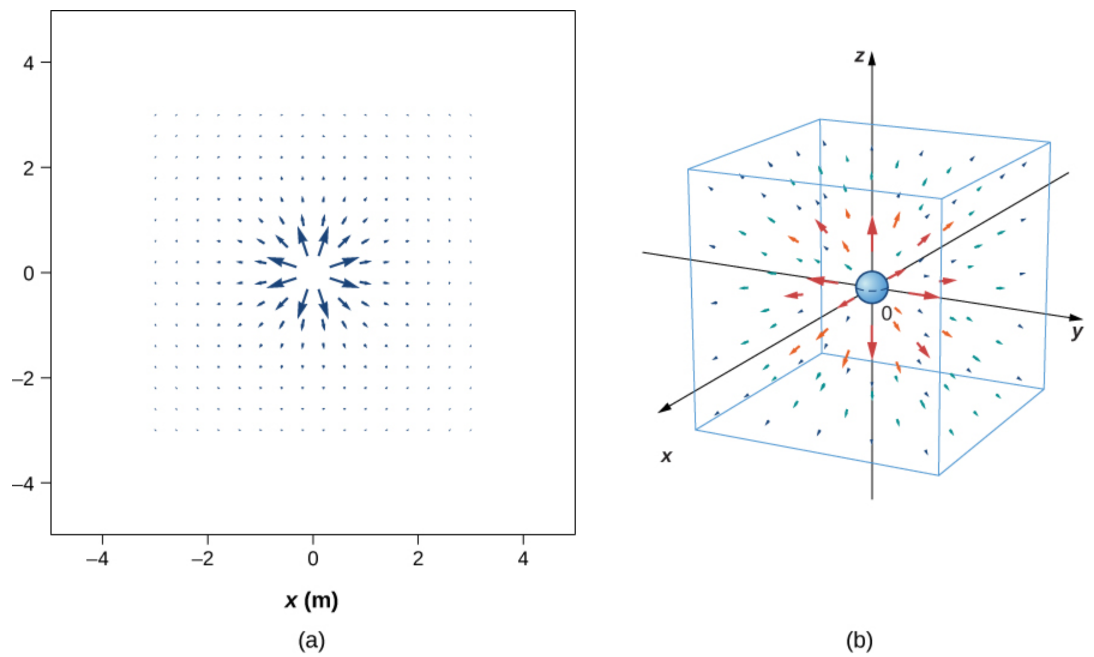


Figure $2.4.1$: The electric field of a positive point charge. A large number of field vectors are shown. Like all vector arrows, the length of each vector is proportional to the magnitude of the field at each point. (a) Field in two dimensions; (b) field in three dimensions.

We've plotted many field vectors in the figure, which are distributed uniformly around the source charge. Since the electric field is a vector, the arrows that we draw correspond at every point in space to both the magnitude and the direction of the field at that point. As always, the length of the arrow that we draw corresponds to the magnitude of the field vector at that point. For a point source charge, the length decreases by the square of the distance from the source charge. In addition, the direction of the field vector is radially away from the source charge, because the direction of the electric field is defined by the direction of the force that a positive test charge would experience in that field. (Again, keep in mind that the actual field is three-dimensional; there are also field lines pointing out of and into the page.)

This diagram is correct, but it becomes less useful as the source charge distribution becomes more complicated. For example, consider the vector field diagram of a dipole (Figure $2.4.2$).

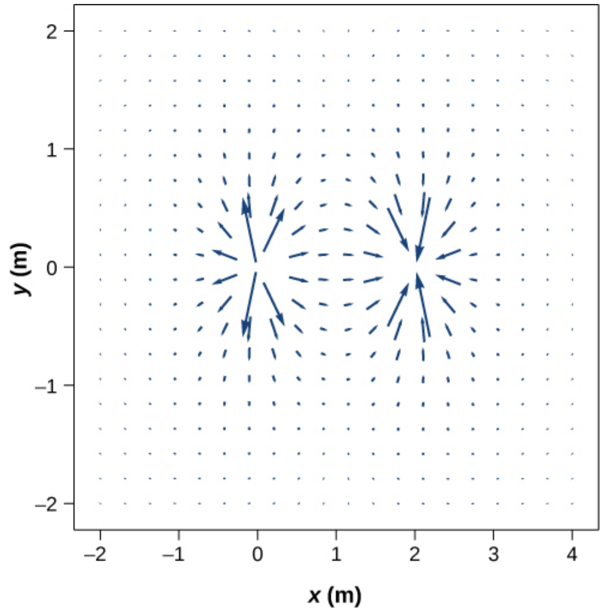


Figure $2.4.2$: The vector field of a dipole. Even with just two identical charges, the vector field diagram becomes difficult to understand.

There is a more useful way to present the same information. Rather than drawing a large number of increasingly smaller vector arrows, we instead connect all of them together, forming continuous lines and curves, as shown in Figure $2.4.3$.

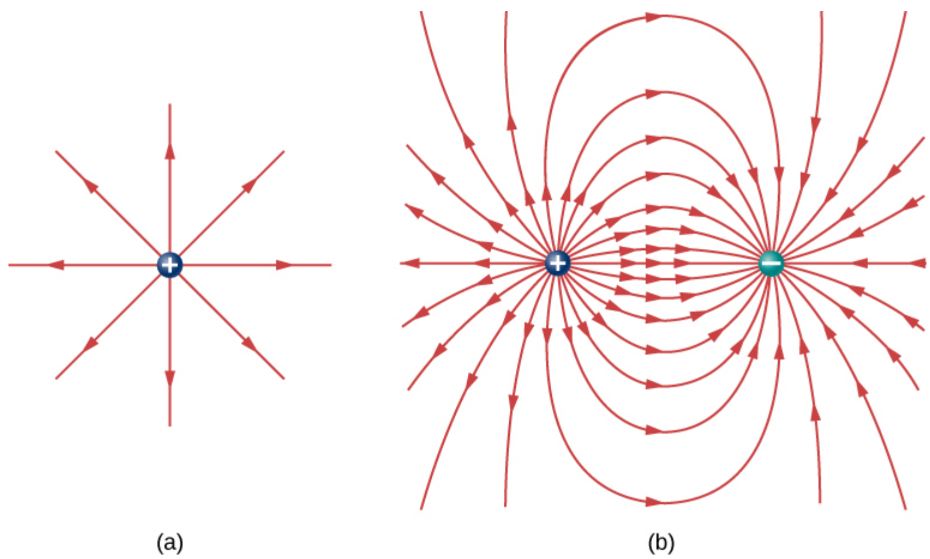


Figure $2.4.3$: (a) The electric field line diagram of a positive point charge. (b) The field line diagram of a dipole. In both diagrams, the magnitude of the field is indicated by the field line density. The field **vectors** (not shown here) are everywhere tangent to the field lines.

Although it may not be obvious at first glance, these field diagrams convey the same information about the electric field as do the vector diagrams. First, the direction of the field at every point is simply the direction of the field vector at that same point. In other words, at any point in space, the field vector at each point is tangent to the field line at that same point. The arrowhead placed on a field line indicates its direction.

As for the magnitude of the field, that is indicated by the **field line density**—that is, the number of field lines per unit area passing through a small cross-sectional area perpendicular to the electric field. This field line density is drawn to be proportional to the magnitude of the field at that cross-section. As a result, if the field lines are close together (that is, the field line density is greater), this indicates that the magnitude of the field is large at that point. If the field lines are far apart at the cross-section, this indicates the magnitude of the field is small. Figure $2.4.4$ shows the idea.

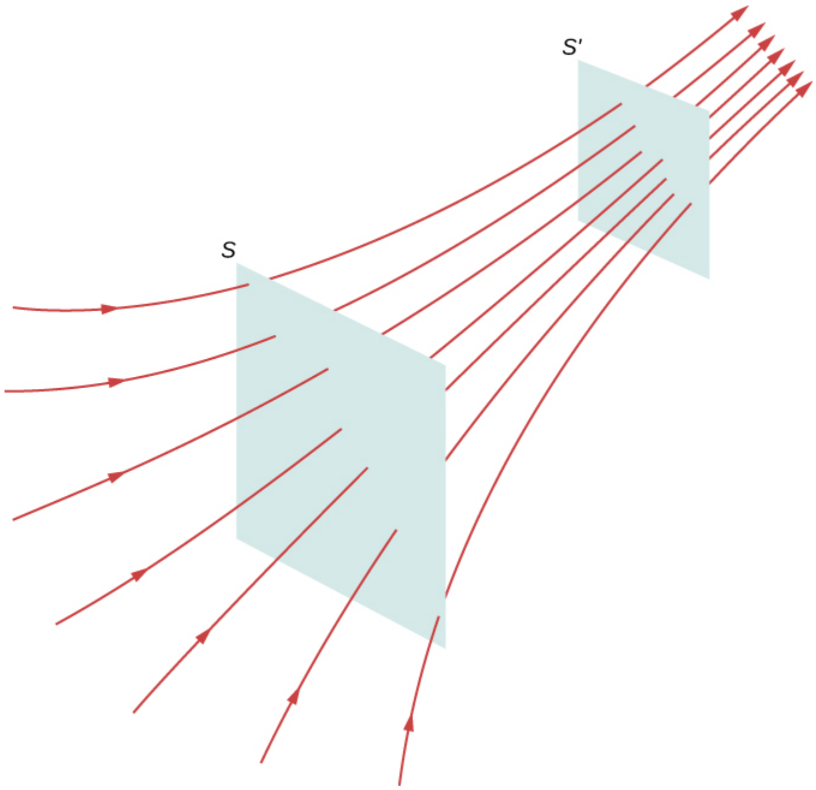


Figure $2.4.4$: Electric field lines passing through imaginary areas. Since the number of lines passing through each area is the same, but the areas themselves are different, the field line density is different. This indicates different magnitudes of the electric field at these points.

In Figure $2.4.4$, the same number of field lines passes through both surfaces S and S', but the surface S is larger than surface S'. Therefore, the density of field lines (number of lines per unit area) is larger at the location of S', indicating that the electric field is stronger at the location of S' than at S. The rules for creating an electric field diagram are as follows.

Problem-Solving Strategy: Drawing Electric Field Lines

1. Electric field lines either originate on positive charges or come in from infinity, and either terminate on negative charges or extend out to infinity.
2. The number of field lines originating or terminating at a charge is proportional to the magnitude of that charge. A charge of $2q$ will have twice as many lines as a charge of q.
3. At every point in space, the field vector at that point is tangent to the field line at that same point.
4. The field line density at any point in space is proportional to (and therefore is representative of) the magnitude of the field at that point in space.
5. Field lines can never cross. Since a field line represents the direction of the field at a given point, if two field lines crossed at some point, that would imply that the electric field was pointing in two different directions at a single point. This in turn would suggest that the (net) force on a test charge placed at that point would point in two different directions. Since this is obviously impossible, it follows that field lines must never cross.

Always keep in mind that field lines serve only as a convenient way to visualize the electric field; they are not physical entities. Although the direction and relative intensity of the electric field can be deduced from a set of field lines, the lines can also be misleading. For example, the field lines drawn to represent the electric field in a region must, by necessity, be discrete. However, the actual electric field in that region exists at every point in space.

Field lines for three groups of discrete charges are shown in Figure $2.4.5$. Since the charges in parts (a) and (b) have the same magnitude, the same number of field lines are shown starting from or terminating on each charge. In (c), however, we draw three times as many field lines leaving the $+3q$ charge as entering the $-q$. The field lines that do not terminate at $-q$ emanate outward from the charge configuration, to infinity.

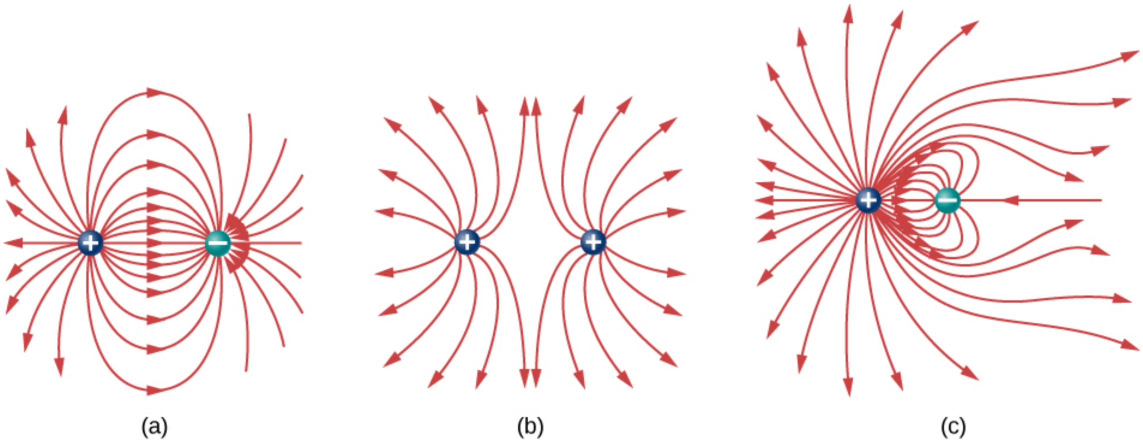


Figure 2.4.5: Three typical electric field diagrams. (a) A dipole. (b) Two identical charges. (c) Two charges with opposite signs and different magnitudes. Can you tell from the diagram which charge has the larger magnitude?

The ability to construct an accurate electric field diagram is an important, useful skill; it makes it much easier to estimate, predict, and therefore calculate the electric field of a source charge. The best way to develop this skill is with software that allows you to place source charges and then will draw the net field upon request. We strongly urge you to search the Internet for a program. Once you've found one you like, run several simulations to get the essential ideas of field diagram construction. Then practice drawing field diagrams, and checking your predictions with the computer-drawn diagrams.

PhET: Charges and Fields

Arrange positive and negative charges in space and view the resulting electric field and electrostatic potential. Plot equipotential lines and discover their relationship to the electric field. Create models of dipoles, capacitors, and more!

CHAPTER OVERVIEW

3: Electric Potentials

3.1: Prelude to Electric Potential

In Electric Charges and Fields, we just scratched the surface (or at least rubbed it) of electrical phenomena. Two terms commonly used to describe electricity are its energy and voltage, which we show in this chapter is directly related to the potential energy in a system. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country via currents through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at the molecular level, ions cross cell membranes and transfer information.

Figure 3.1.1: The energy released in a lightning strike is an excellent illustration of the vast quantities of energy that may be stored and released by an electric potential difference. In this chapter, we calculate just how much energy can be released in a lightning strike and how this varies with the height of the clouds from the ground. (credit: Anthony Quintano)

We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home frequently produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. A motorcycle battery, for example, is small and would not be very successful in replacing a much larger car battery, yet each has the same voltage. In this chapter, we examine the relationship between voltage and electrical energy, and begin to explore some of the many applications of electricity.

3.2: Electric Potential Energy

Learning Objectives

By the end of this section, you will be able to:

- Define the work done by an electric force
- Define electric potential energy
- Apply work and potential energy in systems with electric charges

When a free positive charge q is accelerated by an electric field, it is given kinetic energy (Figure $3.2.1$). The process is analogous to an object being accelerated by a gravitational field, as if the charge were going down an electrical hill where its electric potential energy is converted into kinetic energy, although of course the sources of the forces are very different. Let us explore the work done on a charge q by the electric field in this process, so that we may develop a definition of electric potential energy.

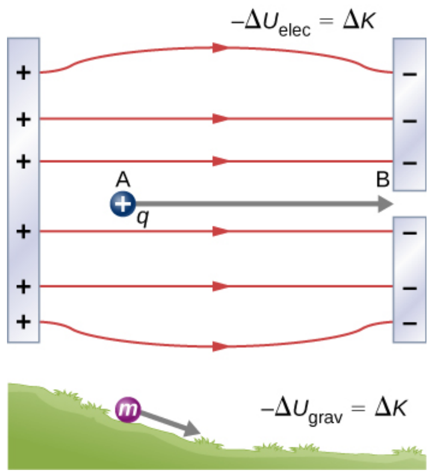


Figure $3.2.1$: A charge accelerated by an electric field is analogous to a mass going down a hill. In both cases, potential energy decreases as kinetic energy increases, $-\Delta U = \Delta K$. Work is done by a force, but since this force is conservative, we can write $W = -\Delta U$.

The electrostatic or Coulomb force is conservative, which means that the work done on q is independent of the path taken, as we will demonstrate later. This is exactly analogous to the gravitational force. When a force is conservative, it is possible to define a potential energy associated with the force. It is usually easier to work with the potential energy (because it depends only on position) than to calculate the work directly.

To show this explicitly, consider an electric charge $+q$ fixed at the origin and move another charge $+Q$ toward q in such a manner that, at each instant, the applied force $\vec{F}$ exactly balances the electric force $\vec{F}_e$ on Q (Figure $3.2.2$). The work done by the applied force $\vec{F}$ on the charge Q changes the potential energy of Q. We call this potential energy the **electrical potential energy** of Q.

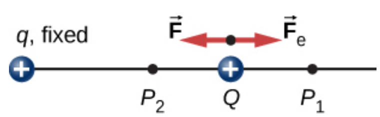


Figure $3.2.2$: Displacement of "test" charge Q in the presence of fixed "source" charge q.

The work W_{12} done by the applied force $\vec{F}$ when the particle moves from P_1 to P_2 may be calculated by

$$W_{12} = \int_{P_1}^{P_2} \vec{F} \cdot d\vec{l}.$$

Since the applied force $\vec{F}$ balances the electric force $\vec{F}_e$ on Q, the two forces have equal magnitude and opposite directions. Therefore, the applied force is

$$\vec{F} = -\vec{F}_e = -\frac{kqQ}{r^2}\hat{r},$$

where we have defined positive to be pointing away from the origin and r is the distance from the origin. The directions of both the displacement and the applied force in the system in Figure $3.2.2$ are parallel, and thus the work done on the system is positive.

We use the letter U to denote electric potential energy, which has units of joules (J). When a conservative force does negative work, the system gains potential energy. When a conservative force does positive work, the system loses potential energy, $\Delta U = -W$. In the system in Figure $3.2.2$, the Coulomb force acts in the opposite direction to the displacement; therefore, the work is negative. However, we have increased the potential energy in the two-charge system.

✔ Example 3.2.1: Kinetic Energy of a Charged Particle

A $+3.0$ nC charge Q is initially at rest a distance of 10 cm (r_1) from a $+5.0$ nC charge q fixed at the origin (Figure $3.2.3$). Naturally, the Coulomb force accelerates Q away from q, eventually reaching 15 cm (r_2).

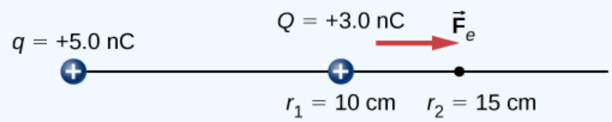


Figure $3.2.3$: The charge Q is repelled by q, thus having work done on it and gaining kinetic energy.

a. What is the work done by the electric field between r_1 and r_2?
b. How much kinetic energy does Q have at r_2?

Strategy

Calculate the work with the usual definition. Since Q started from rest, this is the same as the kinetic energy.

Solution

Integrating force over distance, we obtain

$$\begin{aligned} W_{12} &= \int_{r_1}^{r_2} \vec{F} \cdot d\vec{r} \\ &= \int_{r_1}^{r_2} \frac{kqQ}{r^2} dr \\ &= -\frac{kqQ}{r}\bigg|_{r_1}^{r_2} \\ &= kqQ\left[\frac{-1}{r_2} + \frac{1}{r_1}\right] \\ &= (8.99 \times 10^9\,\mathrm{N\,m^2/C^2})(5.0 \times 10^{-9}\,\mathrm{C})(3.0 \times 10^{-9}\,\mathrm{C})\left[\frac{-1}{0.15\,\mathrm{m}} + \frac{1}{0.10\,\mathrm{m}}\right] \\ &= 4.5 \times 10^{-7}\,\mathrm{J}. \end{aligned}$$

This is also the value of the kinetic energy at r_2.

Significance

Charge Q was initially at rest; the electric field of q did work on Q, so now Q has kinetic energy equal to the work done by the electric field.

? Exercise 3.2.1

If Q has a mass of $4.00\,\mu\mathrm{g}$, what is the speed of Q at r_2?

Answer

$$K = \frac{1}{2}mv^2,\ v = \sqrt{2\frac{\mathrm{K}}{\mathrm{m}}} = \sqrt{2\frac{4.5\times10^{-7}\mathrm{J}}{4.00\times10^{-9}\mathrm{kg}}} = 15\,\mathrm{m/s}.$$

In this example, the work W done to accelerate a positive charge from rest is positive and results from a loss in U, or a negative ΔU. A value for U can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Electric Potential Energy

Work W done to accelerate a positive charge from rest is positive and results from a loss in U, or a negative ΔU. Mathematically,

$$W = -\Delta U. \tag{3.2.1}$$

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of electric potential energy than to deal with the Coulomb force directly in real-world applications.

In polar coordinates with q at the origin and Q located at r, the displacement element vector is $\vec{dl} = \hat{r}\,dr$ and thus the work becomes

$$\begin{aligned} W_{12} &= kqQ \int_{r_1}^{r_2} \frac{1}{r^2}\hat{r}\cdot\hat{r}\,dr \\ &= \underbrace{kqQ\frac{1}{r_2}}_{\text{final point}} - \underbrace{kqQ\frac{1}{r_1}}_{\text{initial point}}. \end{aligned} \tag{3.2.2}$$

Notice that this result only depends on the endpoints and is otherwise independent of the path taken. To explore this further, compare path P_1 to P_2 with path $P_1P_3P_4P_2$ in Figure $3.2.4$.

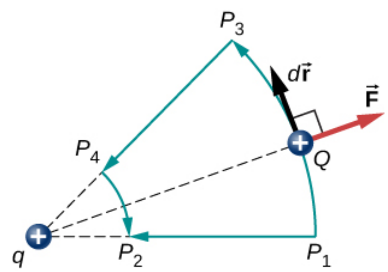


Figure $3.2.4$:Two paths for displacement P_1 to P_2. The work on segments P_1P_3 and P_4P_2 are zero due to the electrical force being perpendicular to the displacement along these paths. Therefore, work on paths P_1P_2 and $P_1P_3P_4P_2$ are equal.

The segments P_1P_3 and P_4P_2 are arcs of circles centered at q. Since the force on Q points either toward or away from q, no work is done by a force balancing the electric force, because it is perpendicular to the displacement along these arcs. Therefore, the only work done is along segment P_3P_4 which is identical to P_1P_2.

One implication of this work calculation is that if we were to go around the path $P_1P_3P_4P_2P_1$, the net work would be zero (Figure $3.2.5$). Recall that this is how we determine whether a force is conservative or not. Hence, because the electric force is related to the electric field by $\vec{F} = q\vec{E}$, the electric field is itself conservative. That is,

$$\oint \vec{E}\cdot\vec{dl} = 0.$$

Note that Q is a constant.

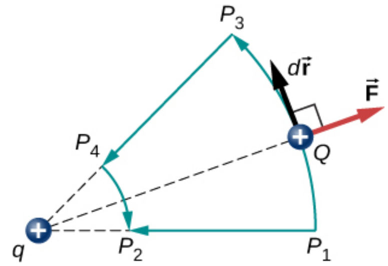


Figure 3.2.5: A closed path in an electric field. The net work around this path is zero.

Another implication is that we may define an electric potential energy. Recall that the work done by a conservative force is also expressed as the difference in the potential energy corresponding to that force. Therefore, the work W_{ref} to bring a charge from a reference point to a point of interest may be written as

$$W_{ref} = \int_{r_{\text{ref}}}^{r} \vec{F} \cdot d\vec{l}$$

and, by Equation 3.2.1, the difference in potential energy ($U_2 - U_1$) of the test charge Q between the two points is

$$\Delta U = -\int_{r_{\text{ref}}}^{r} \vec{F} \cdot d\vec{l}.$$

Therefore, we can write a general expression for the potential energy of two point charges (in spherical coordinates):

$$\Delta U = -\int_{r_{\text{ref}}}^{r} \frac{kqQ}{r^2} dr = -\left[-\frac{kqQ}{r}\right]_{r_{\text{ref}}}^{r} = kqQ\left[\frac{1}{r} - \frac{1}{r_{\text{ref}}}\right].$$

We may take the second term to be an arbitrary constant reference level, which serves as the zero reference:

$$U(r) = k\frac{qQ}{r} - U_{\text{ref}}.$$

A convenient choice of reference that relies on our common sense is that when the two charges are infinitely far apart, there is no interaction between them. (Recall the discussion of reference potential energy in Potential Energy and Conservation of Energy.) Taking the potential energy of this state to be zero removes the term U_{ref} from the equation (just like when we say the ground is zero potential energy in a gravitational potential energy problem), and the potential energy of Q when it is separated from q by a distance r assumes the form

$$\underbrace{U(r) = k\frac{qQ}{r}}_{\text{zero reference at } r=\infty}.$$

This formula is symmetrical with respect to q and Q, so it is best described as the potential energy of the two-charge system.

✔ Example 3.2.2: Potential Energy of a Charged Particle

A $+3.0\text{ nC}$ charge Q is initially at rest a distance of 10 cm (r_1) from a $+5.0\text{ nC}$ charge q fixed at the origin (Figure 3.2.6). Naturally, the Coulomb force accelerates Q away from q, eventually reaching 15 cm (r_2).

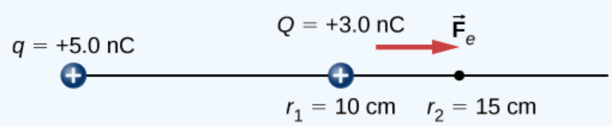


Figure 3.2.6: The charge Q is repelled by q, thus having work done on it and losing potential energy.

What is the change in the potential energy of the two-charge system from r_1 to r_2?

Strategy

Calculate the potential energy with the definition given above:

$\Delta U_{12} = -\int_{r_1}^{r_2} \vec{F} \cdot d\vec{r}$. Since Q started from rest, this is the same as the kinetic energy.

Solution

We have

$$\begin{aligned}\Delta U_{12} &= -\int_{r_1}^{r_2} \vec{F}\cdot d\vec{r} \\ &= -\int_{r_1}^{r_2} \frac{kqQ}{r^2}dr \\ &= -\left[-\frac{kqQ}{r}\right]_{r_1}^{r_2} \\ &= kqQ\left[\frac{1}{r_2}-\frac{1}{r_1}\right] \\ &= (8.99\times 10^9\,\mathrm{N\,m^2/C^2})(5.0\times 10^{-9}\,\mathrm{C})(3.0\times 10^{-9}\,\mathrm{C})\left[\frac{1}{0.15\,\mathrm{m}}-\frac{1}{0.10\,\mathrm{m}}\right] \\ &= -4.5\times 10^{-7}\,\mathrm{J}.\end{aligned} \tag{3.2.3}$$

Significance

The change in the potential energy is negative, as expected, and equal in magnitude to the change in kinetic energy in this system. Recall from Example $3.2.1$ that the change in kinetic energy was positive.

Exercise $3.2.2$

What is the potential energy of Q relative to the zero reference at infinity at r_2 in the above example?

Answer

It has kinetic energy of $4.5\times 10^{-7}\,\mathrm{J}$ at point r_2 and potential energy of $9.0\times 10^{-7}\,\mathrm{J}$, which means that as Q approaches infinity, its kinetic energy totals three times the kinetic energy at r_2, since all of the potential energy gets converted to kinetic.

Due to Coulomb's law, the forces due to multiple charges on a test charge Q superimpose; they may be calculated individually and then added. This implies that the work integrals and hence the resulting potential energies exhibit the same behavior. To demonstrate this, we consider an example of assembling a system of four charges.

Example $3.2.3$: Assembling Four Positive Charges

Find the amount of work an external agent must do in assembling four charges $+2.0\,\mu$C, $+3.0\,\mu$C, $+4.0\,\mu$C and $+5.0\,\mu$C at the vertices of a square of side 1.0 cm, starting each charge from infinity (Figure $3.2.7$).

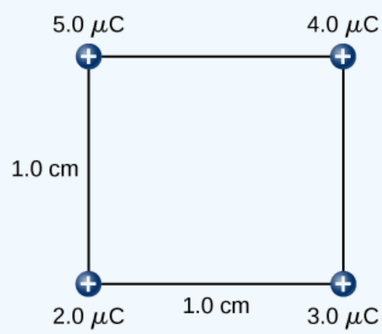


Figure $3.2.7$: How much work is needed to assemble this charge configuration?

Strategy

We bring in the charges one at a time, giving them starting locations at infinity and calculating the work to bring them in from infinity to their final location. We do this in order of increasing charge.

Solution

Step 1. First bring the $+2.0\ \mu\text{C}$ charge to the origin. Since there are no other charges at a finite distance from this charge yet, no work is done in bringing it from infinity,

$$W_1 = 0.$$

Step 2. While keeping the $+2.0\ \mu\text{C}$ charge fixed at the origin, bring the $+3.0\ \mu\text{C}$ charge to $(x, y, z) = (1.0\ \text{cm},\ 0,\ 0)$ (Figure 3.2.8). Now, the applied force must do work against the force exerted by the $+2.0\ \mu\text{C}$ charge fixed at the origin. The work done equals the change in the potential energy of the $+3.0\ \mu\text{C}$ charge:

$$\begin{aligned} W_2 &= k\frac{q_1 q_2}{r12} \\ &= \left(9.0\times 10^9\,\frac{\text{N}\cdot\text{m}^2}{\text{C}^2}\right)\frac{(2.0\times 10^{-6}\,\text{C})(3.0\times 10^{-6}\,\text{C})}{1.0\times 10^{-2}\,\text{m}} \\ &= 5.4\ \text{J}. \end{aligned}$$

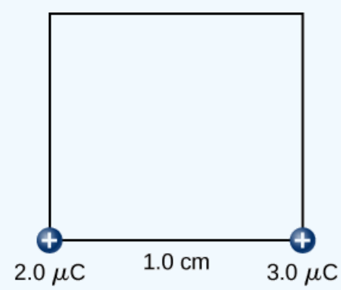


Figure 3.2.8: Step 2. Work W_2 to bring the $+3.0\ \mu\text{C}$ charge from infinity.

Step 3. While keeping the charges of $+2.0\ \mu\text{C}$ and $+3.0\ \mu\text{C}$ fixed in their places, bring in the $+4.0\ \mu\text{C}$ charge to $(x, y, z) = (1.0\ \text{cm},\ 1.0\ \text{cm},\ 0)$(Figure 3.2.9). The work done in this step is

$$\begin{aligned} W_3 &= k\frac{q_1 q_3}{r_{13}} + k\frac{q_2 q_3}{r_{23}} \\ &= \left(9.0\times 10^9\,\frac{\text{N}\cdot\text{m}^2}{\text{C}^2}\right)\left[\frac{(2.0\times 10^{-6}\,\text{C})(4.0\times 10^{-6}\,\text{C})}{\sqrt{2}\times 10^{-2}\,\text{m}} + \frac{(3.0\times 10^{-6}\,\text{C})(4.0\times 10^{-6}\,\text{C})}{1.0\times 10^{-2}\,\text{m}}\right] \\ &= 15.9\ \text{J}. \end{aligned}$$

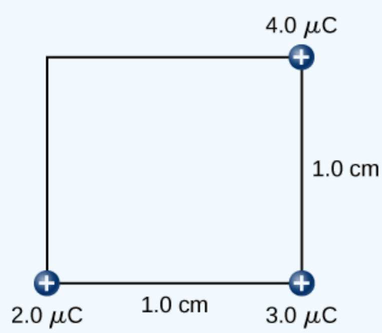


Figure 3.2.9: Step 3. The work W_3 to bring the $+4.0\ \mu\text{C}$ charge from infinity.

Step 4. Finally, while keeping the first three charges in their places, bring the $+5.0\ \mu\text{C}$ charge to $(x, y, z) = (0,\ 1.0\ \text{cm},\ 0)$ (Figure 3.2.10). The work done here is

$$\begin{aligned} W_4 &= kq_4\left[\frac{q_1}{r_{14}} + \frac{q_2}{r_{24}} + \frac{q_3}{r_{34}}\right], \\ &= \left(9.0\times 10^9\,\frac{\text{N}\cdot\text{m}^2}{\text{C}^2}\right)(5.0\times 10^{-6}\,\text{C})\left[\frac{(2.0\times 10^{-6}\,\text{C})}{1.0\times 10^{-2}\,\text{m}} + \frac{(3.0\times 10^{-6}\,\text{C})}{\sqrt{2}\times 10^{-2}\,\text{m}} + \frac{(4.0\times 10^{-6}\,\text{C})}{1.0\times 10^{-2}\,\text{m}}\right] \\ &= 36.5\ \text{J}. \end{aligned}$$

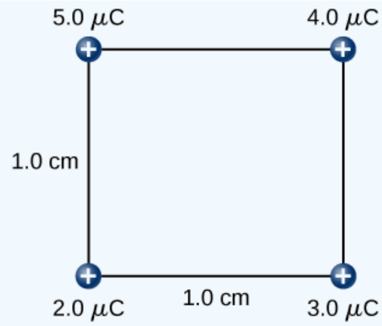


Figure 3.2.10: Step 4. The work W_4 to bring the $+5.0\ \mu\mathrm{C}$ charge from infinity.

Hence, the total work done by the applied force in assembling the four charges is equal to the sum of the work in bringing each charge from infinity to its final position:

$$\begin{aligned} W_T &= W_1 + W_2 + W_3 + W_4 \\ &= 0 + 5.4\ \mathrm{J} + 15.9\ \mathrm{J} + 36.5\ \mathrm{J} \\ &= 57.8\ \mathrm{J}. \end{aligned}$$

Significance

The work on each charge depends only on its pairwise interactions with the other charges. No more complicated interactions need to be considered; the work on the third charge only depends on its interaction with the first and second charges, the interaction between the first and second charge does not affect the third.

? Exercise $3.2.3$

Is the electrical potential energy of two point charges positive or negative if the charges are of the same sign? Opposite signs? How does this relate to the work necessary to bring the charges into proximity from infinity?

Answer

Positive, negative, and these quantities are the same as the work you would need to do to bring the charges in from infinity

Note that the electrical potential energy is positive if the two charges are of the same type, either positive or negative, and negative if the two charges are of opposite types. This makes sense if you think of the change in the potential energy ΔU as you bring the two charges closer or move them farther apart. Depending on the relative types of charges, you may have to work on the system or the system would do work on you, that is, your work is either positive or negative. If you have to do positive work on the system (actually push the charges closer), then the energy of the system should increase. If you bring two positive charges or two negative charges closer, you have to do positive work on the system, which raises their potential energy. Since potential energy is proportional to $1/r$, the potential energy goes up when r goes down between two positive or two negative charges.

On the other hand, if you bring a positive and a negative charge nearer, you have to do negative work on the system (the charges are pulling you), which means that you take energy away from the system. This reduces the potential energy. Since potential energy is negative in the case of a positive and a negative charge pair, the increase in $1/r$ makes the potential energy more negative, which is the same as a reduction in potential energy.

The result from Example $3.2.2$ may be extended to systems with any arbitrary number of charges. In this case, it is most convenient to write the formula as

$$W_{12\ldots N} = \frac{k}{2} \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{r_{ij}} \text{ for } i \neq j.$$

The factor of 1/2 accounts for adding each pair of charges twice.

- **7.2: Electric Potential Energy** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/university-physics-volume-2.

3.3: Electric Potential and Potential Difference

Learning Objectives

By the end of this section, you will be able to:

- Define electric potential, voltage, and potential difference
- Define the electron-volt
- Calculate electric potential and potential difference from potential energy and electric field
- Describe systems in which the electron-volt is a useful unit
- Apply conservation of energy to electric systems

Recall that earlier we defined electric field to be a quantity independent of the test charge in a given system, which would nonetheless allow us to calculate the force that would result on an arbitrary test charge. (The default assumption in the absence of other information is that the test charge is positive.) We briefly defined a field for gravity, but gravity is always attractive, whereas the electric force can be either attractive or repulsive. Therefore, although potential energy is perfectly adequate in a gravitational system, it is convenient to define a quantity that allows us to calculate the work on a charge independent of the magnitude of the charge. Calculating the work directly may be difficult, since $W = \vec{F} \cdot \vec{d}$ and the direction and magnitude of $\vec{F}$ can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that because $\vec{F}$, the work, and hence ΔU is proportional to the test charge q. To have a physical quantity that is independent of test charge, we define **electric potential** V (or simply potential, since electric is understood) to be the potential energy per unit charge:

Electric Potential

The electric potential energy per unit charge is

$$V = \frac{U}{q}. \tag{3.3.1}$$

Since U is proportional to q, the dependence on q cancels. Thus, V does not depend on q. The change in potential energy ΔU is crucial, so we are concerned with the difference in potential or potential difference ΔV between two points, where

Electric Potential Difference

The **electric potential difference** between points A and B, $V_B - V_A$ is defined to be the change in potential energy of a charge q moved from A to B, divided by the charge. Units of potential difference are joules per coulomb, given the name volt (V) after Alessandro Volta.

$$1\text{ V} = 1\text{ J/C} \tag{3.3.2}$$

The familiar term **voltage** is the common name for electric potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor. It is worthwhile to emphasize the distinction between potential difference and electrical potential energy.

Potential Difference and Electrical Potential Energy

The relationship between potential difference (or voltage) and electrical potential energy is given by

$$\Delta V = \frac{\Delta U}{q} \tag{3.3.3}$$

or

$$\Delta U = q\Delta V. \tag{3.3.4}$$

Voltage is not the same as energy. Voltage is the energy per unit charge. Thus, a motorcycle battery and a car battery can both have the same voltage (more precisely, the same potential difference between battery terminals), yet one stores much more energy than the other because $\Delta U = q\Delta V$. The car battery can move more charge than the motorcycle battery, although both are 12-V batteries.

Example 3.3.1: Calculating Energy

You have a 12.0-V motorcycle battery that can move 5000 C of charge, and a 12.0-V car battery that can move 60,000 C of charge. How much energy does each deliver? (Assume that the numerical value of each charge is accurate to three significant figures.)

Strategy

To say we have a 12.0-V battery means that its terminals have a 12.0-V potential difference. When such a battery moves charge, it puts the charge through a potential difference of 12.0 V, and the charge is given a change in potential energy equal to $\Delta U = q\Delta V$. To find the energy output, we multiply the charge moved by the potential difference.

Solution

For the motorcycle battery, $q = 5000\,\mathrm{C}$ and $\Delta V = 12.0\,\mathrm{V}$. The total energy delivered by the motorcycle battery is

$$\Delta U_{cycle} = (5000\,\mathrm{C})(12.0\,\mathrm{V}) = (5000\,\mathrm{C})(12.0\,\mathrm{J/C}) = 6.00 \times 10^4\,\mathrm{J}.$$

Similarly, for the car battery, $q = 60,000\,\mathrm{C}$ and

$$\Delta U_{car} = (60,000\,\mathrm{C})(12.0\,\mathrm{V}) = 7.20 \times 10^5\,\mathrm{J}.$$

Significance

Voltage and energy are related, but they are not the same thing. The voltages of the batteries are identical, but the energy supplied by each is quite different. A car battery has a much larger engine to start than a motorcycle. Note also that as a battery is discharged, some of its energy is used internally and its terminal voltage drops, such as when headlights dim because of a depleted car battery. The energy supplied by the battery is still calculated as in this example, but not all of the energy is available for external use.

Exercise 3.3.1

How much energy does a 1.5-V AAA battery have that can move 100 C?

Answer

$$\Delta U = q\Delta V = (100\,\mathrm{C})(1.5\,\mathrm{V}) = 150\,\mathrm{J}$$

Note that the energies calculated in the previous example are absolute values. The change in potential energy for the battery is negative, since it loses energy. These batteries, like many electrical systems, actually move negative charge—electrons in particular. The batteries repel electrons from their negative terminals (A) through whatever circuitry is involved and attract them to their positive terminals (B), as shown in Figure $3.3.1$. The change in potential is $\Delta V = V_B - V_A = +12\,V$ and the charge q is negative, so that $\Delta U = q\Delta V$ is negative, meaning the potential energy of the battery has decreased when q has moved from A to B.

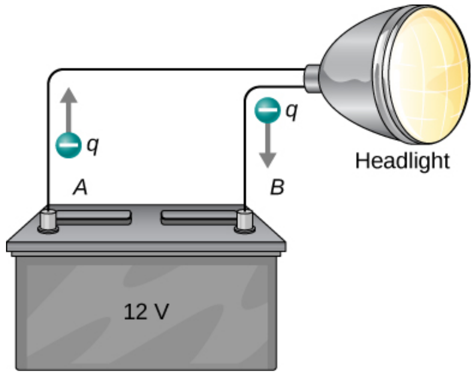


Figure 3.3.1: A battery moves negative charge from its negative terminal through a headlight to its positive terminal. Appropriate combinations of chemicals in the battery separate charges so that the negative terminal has an excess of negative charge, which is repelled by it and attracted to the excess positive charge on the other terminal. In terms of potential, the positive terminal is at a higher voltage than the negative terminal. Inside the battery, both positive and negative charges move.

✔ Example 3.3.2: How Many Electrons Move through a Headlight Each Second?

When a 12.0-V car battery powers a single 30.0-W headlight, how many electrons pass through it each second?

Strategy

To find the number of electrons, we must first find the charge that moves in 1.00 s. The charge moved is related to voltage and energy through the equations $\Delta U = q\Delta V$. A 30.0-W lamp uses 30.0 joules per second. Since the battery loses energy, we have $\Delta U = -30$ J and, since the electrons are going from the negative terminal to the positive, we see that $\Delta V = +12.0\,\text{V}$.

Solution

To find the charge q moved, we solve the equation $\Delta U = q\Delta V$:

$$q = \frac{\Delta U}{\Delta V}.$$

Entering the values for ΔU and ΔV, we get

$$q = \frac{-30.0\,\text{J}}{+12.0\,\text{V}} = \frac{-30.0\,\text{J}}{+12.0\,\text{J/C}} = -2.50\,\text{C}.$$

The number of electrons n_e is the total charge divided by the charge per electron. That is,

$$n_e = \frac{-2.50\,\text{C}}{-1.60\times 10^{-19}\text{C}/e^-} = 1.56\times 10^{19}\,\text{electrons}.$$

Significance

This is a very large number. It is no wonder that we do not ordinarily observe individual electrons with so many being present in ordinary systems. In fact, electricity had been in use for many decades before it was determined that the moving charges in many circumstances were negative. Positive charge moving in the opposite direction of negative charge often produces identical effects; this makes it difficult to determine which is moving or whether both are moving.

? Exercise 3.3.2

How many electrons would go through a 24.0-W lamp?

Answer

$-2.00\,C,\ n_e = 1.25\times 10^{19}\,electrons$

The Electron-Volt

The energy per electron is very small in macroscopic situations like that in the previous example—a tiny fraction of a joule. But on a submicroscopic scale, such energy per particle (electron, proton, or ion) can be of great importance. For example, even a tiny fraction of a joule can be great enough for these particles to destroy organic molecules and harm living tissue. The particle may do its damage by direct collision, or it may create harmful X-rays, which can also inflict damage. It is useful to have an energy unit related to submicroscopic effects.

Figure $3.3.2$ shows a situation related to the definition of such an energy unit. An electron is accelerated between two charged metal plates, as it might be in an old-model television tube or oscilloscope. The electron gains kinetic energy that is later converted into another form—light in the television tube, for example. (Note that in terms of energy, "downhill" for the electron is "uphill" for a positive charge.) Since energy is related to voltage by $\Delta U = q\Delta V$, we can think of the joule as a coulomb-volt.

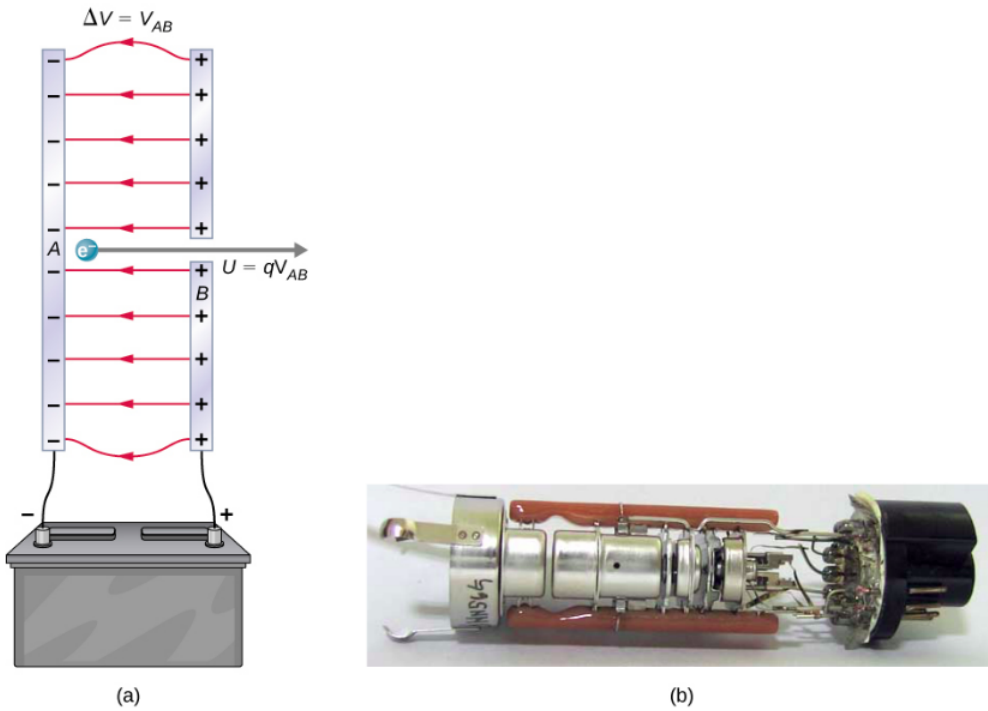


Figure $3.3.2$: A typical electron gun accelerates electrons using a potential difference between two separated metal plates. By conservation of energy, the kinetic energy has to equal the change in potential energy, so $KE = qV$. The energy of the electron in electron-volts is numerically the same as the voltage between the plates. For example, a 5000-V potential difference produces 5000-eV electrons. The conceptual construct, namely two parallel plates with a hole in one, is shown in (a), while a real electron gun is shown in (b).

The Electron-Volt Unit

On the submicroscopic scale, it is more convenient to define an energy unit called the **electron-volt** (eV), which is the energy given to a fundamental charge accelerated through a potential difference of 1 V. In equation form,

$$1\,\text{eV} = (1.60 \times 10^{-19}\,\text{C})(1\,\text{V}) = (1.60 \times 10^{-19}\,\text{C})(1\,\text{J/C}) = 1.60 \times 10^{-19}\,\text{J}.$$

An electron accelerated through a potential difference of 1 V is given an energy of 1 eV. It follows that an electron accelerated through 50 V gains 50 eV. A potential difference of 100,000 V (100 kV) gives an electron an energy of 100,000 eV (100 keV), and so on. Similarly, an ion with a double positive charge accelerated through 100 V gains 200 eV of energy. These simple relationships between accelerating voltage and particle charges make the electron-volt a simple and convenient energy unit in such circumstances.

The electron-volt is commonly employed in submicroscopic processes—chemical valence energies and molecular and nuclear binding energies are among the quantities often expressed in electron-volts. For example, about 5 eV of energy is required to break up certain organic molecules. If a proton is accelerated from rest through a potential difference of 30 kV, it acquires an energy of 30 keV (30,000 eV) and can break up as many as 6000 of these molecules $(30,000\,\text{eV} \div 5\,\text{eV per molecule} = 6000\,\text{molecules})$ Nuclear decay energies are on the order of 1 MeV (1,000,000 eV) per event and can thus produce significant biological damage.

Conservation of Energy

The total energy of a system is conserved if there is no net addition (or subtraction) due to work or heat transfer. For conservative forces, such as the electrostatic force, conservation of energy states that mechanical energy is a constant.

Mechanical energy is the sum of the kinetic energy and potential energy of a system; that is, $K+U=\text{constant}$. A loss of (\U\) for a charged particle becomes an increase in its (\K\). Conservation of energy is stated in equation form as

$$K+U=\text{constant}$$

or

$$K_i+U_i=K_f+U_f$$

where i and f stand for initial and final conditions. As we have found many times before, considering energy can give us insights and facilitate problem solving.

Example 3.3.3: Electrical Potential Energy Converted into Kinetic Energy

Calculate the final speed of a free electron accelerated from rest through a potential difference of 100 V. (Assume that this numerical value is accurate to three significant figures.)

Strategy

We have a system with only conservative forces. Assuming the electron is accelerated in a vacuum, and neglecting the gravitational force (we will check on this assumption later), all of the electrical potential energy is converted into kinetic energy. We can identify the initial and final forms of energy to be

$K_i=0, K_f=\frac{1}{2}mv^2, U_i=qV, U_f=0.$

Solution

Conservation of energy states that

$$K_i+U_i=K_f+U_f.$$

Entering the forms identified above, we obtain

$$qV=\frac{mv^2}{2}.$$

We solve this for (\v\):

$$v=\sqrt{\frac{2qV}{m}}.$$

Entering values for (\q\), (\V\), and (\m\) gives

$$v=\sqrt{\frac{2(-1.60\times 10^{-19}\text{C})(-100\text{ J/C})}{9.11\times 10^{-31}\text{kg}}}=5.93\times 10^6\text{ m/s}.$$

Significance

Note that both the charge and the initial voltage are negative, as in Figure 3.3.2. From the discussion of electric charge and electric field, we know that electrostatic forces on small particles are generally very large compared with the gravitational force. The large final speed confirms that the gravitational force is indeed negligible here. The large speed also indicates how easy it is to accelerate electrons with small voltages because of their very small mass. Voltages much higher than the 100 V in this problem are typically used in electron guns. These higher voltages produce electron speeds so great that effects from special relativity must be taken into account and will be discussed elsewhere. That is why we consider a low voltage (accurately) in this example.

? Exercise 3.3.3

How would this example change with a positron? A positron is identical to an electron except the charge is positive.

Answer

It would be going in the opposite direction, with no effect on the calculations as presented.

Voltage and Electric Field

So far, we have explored the relationship between voltage and energy. Now we want to explore the relationship between voltage and electric field. We will start with the general case for a non-uniform $\vec{E}$ field. Recall that our general formula for the potential energy of a test charge q at point P relative to reference point R is

$$U_p = -\int_R^p \vec{F} \cdot d\vec{l}.$$

When we substitute in the definition of electric field $(\vec{E} = \vec{F}/q)$, this becomes

$$U_p = -q\int_R^p \vec{E} \cdot d\vec{l}.$$

Applying our definition of potential $(V = U/q)$ to this potential energy, we find that, in general,

$$V_p = -\int_R^p \vec{E} \cdot d\vec{l}.$$

From our previous discussion of the potential energy of a charge in an electric field, the result is independent of the path chosen, and hence we can pick the integral path that is most convenient.

Consider the special case of a positive point charge q at the origin. To calculate the potential caused by q at a distance r from the origin relative to a reference of 0 at infinity (recall that we did the same for potential energy), let $P = r$ and $R = \infty$, with $d\vec{l} = d\vec{r} = \hat{r}\,dr$ and use $\vec{E} = \frac{kq}{r^2}\hat{r}$. When we evaluate the integral

$$V_p = -\int_R^p \vec{E} \cdot d\vec{l}$$

for this system, we have

$$V_r = -\int_\infty^r \frac{kq}{r^2} dr = \frac{kq}{r} - \frac{kq}{\infty} = \frac{kq}{r}.$$

This result,

$$V_r = \frac{kq}{r}$$

is the standard form of the potential of a point charge. This will be explored further in the next section.

To examine another interesting special case, suppose a uniform electric field $\vec{E}$ is produced by placing a potential difference (or voltage) ΔV across two parallel metal plates, labeled A and B (Figure $3.3.3$). Examining this situation will tell us what voltage is needed to produce a certain electric field strength. It will also reveal a more fundamental relationship between electric potential and electric field.

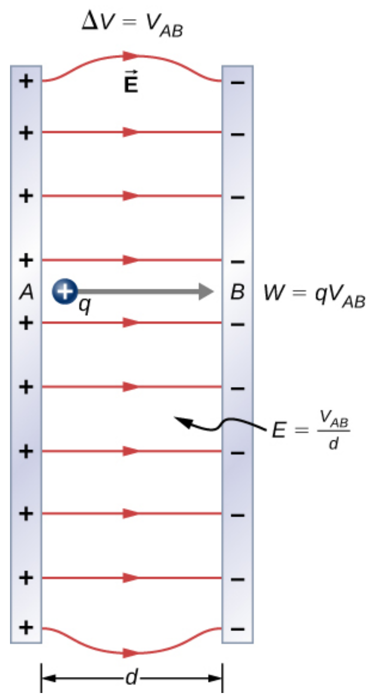


Figure 3.3.3: The relationship between V and E for parallel conducting plates is $E = V/d$. (Note that $\Delta V = V_{AB}$ in magnitude. For a charge that is moved from plate A at higher potential to plate B at lower potential, a minus sign needs to be included as follows: $-\Delta V = V_A - V_B = V_{AB}$.)

From a physicist's point of view, either ΔV or $\vec{E}$ can be used to describe any interaction between charges. However, ΔV is a scalar quantity and has no direction, whereas $\vec{E}$ is a vector quantity, having both magnitude and direction. (Note that the magnitude of the electric field, a scalar quantity, is represented by E.) The relationship between ΔV and $\vec{E}$ is revealed by calculating the work done by the electric force in moving a charge from point A to point B. But, as noted earlier, arbitrary charge distributions require calculus. We therefore look at a uniform electric field as an interesting special case.

The work done by the electric field in Figure $3.3.3$ to move a positive charge q from A, the positive plate, higher potential, to B, the negative plate, lower potential, is

$$W = -\Delta U = -q\Delta V.$$

The potential difference between points A and B is

$$-\Delta V = -(V_B - V_A) = V_A - V_B = V_{AB}.$$

Entering this into the expression for work yields

$$W = qV_{AB}.$$

Work is $W = \vec{F} \cdot \vec{d} = Fd\cos\theta$: here $\cos\theta = 1$, since the path is parallel to the field. Thus, $W = Fd$. Since $F = qE$ we see that $W = qEd$.

Substituting this expression for work into the previous equation gives

$$qEd = qV_{AB}.$$

The charge cancels, so we obtain for the voltage between points A and B.

In uniform E-field only:

$$V_{AB} = Ed$$

$$E = \frac{V_{AB}}{d}$$

where d is the distance from A to B, or the distance between the plates in Figure $3.3.3$. Note that this equation implies that the units for electric field are volts per meter. We already know the units for electric field are newtons per coulomb; thus, the following relation among units is valid:

$$1\,\mathrm{N/C} = 1\,\mathrm{V/m}.$$

Furthermore, we may extend this to the integral form. Substituting Equation 3.3.3 into our definition for the potential difference between points A and B we obtain

$$V_{AB} = V_B - V_A = -\int_R^B \vec{E}\cdot d\vec{l} + \int_R^A \vec{E}\cdot d\vec{l}$$

which simplifies to

$$V_B - V_A = -\int_A^B \vec{E}\cdot d\vec{l}.$$

As a demonstration, from this we may calculate the potential difference between two points (A and B) equidistant from a point charge q at the origin, as shown in Figure $3.3.4$.

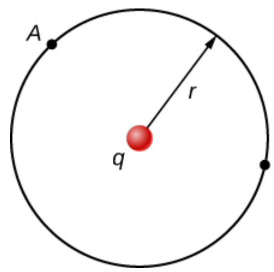


Figure $3.3.4$: The arc for calculating the potential difference between two points that are equidistant from a point charge at the origin.

To do this, we integrate around an arc of the circle of constant radius r between A and B, which means we let $d\vec{l} = r\hat{\varphi}d\varphi$, while using $\vec{E} = \frac{kq}{r^2}\hat{r}$. Thus,

$$\Delta V = V_B - V_A = -\int_A^B \vec{E}\cdot d\vec{l}.$$

for this system becomes

$$V_B - V_A = -\int_A^B \frac{kq}{r^2}\hat{r}\cdot r\hat{\varphi}d\varphi.$$

However, $\hat{r}\cdot\hat{\varphi}$ and therefore

$$V_B - V_A = 0.$$

This result, that there is no difference in potential along a constant radius from a point charge, will come in handy when we map potentials.

Example $3.3.4A$: What Is the Highest Voltage Possible between Two Plates?

Dry air can support a maximum electric field strength of about $3.0\times10^6\ \mathrm{V/m}$. Above that value, the field creates enough ionization in the air to make the air a conductor. This allows a discharge or spark that reduces the field. What, then, is the maximum voltage between two parallel conducting plates separated by 2.5 cm of dry air?

Strategy

We are given the maximum electric field E between the plates and the distance d between them. We can use the equation $V_{AB} = Ed$ to calculate the maximum voltage.

Solution

The potential difference or voltage between the plates is

$$V_{AB} = Ed.$$

Entering the given values for E and d gives

$$V_{AB} = (3.0 \times 10^6\ \mathrm{V/m})(0.025\,\mathrm{m}) = 7.5 \times 10^4\ \mathrm{V}$$

or

$$V_{AB} = 75\,\mathrm{kV}.$$

(The answer is quoted to only two digits, since the maximum field strength is approximate.)

Significance

One of the implications of this result is that it takes about 75 kV to make a spark jump across a 2.5-cm (1-in.) gap, or 150 kV for a 5-cm spark. This limits the voltages that can exist between conductors, perhaps on a power transmission line. A smaller voltage can cause a spark if there are spines on the surface, since sharp points have larger field strengths than smooth surfaces. Humid air breaks down at a lower field strength, meaning that a smaller voltage will make a spark jump through humid air. The largest voltages can be built up with static electricity on dry days (Figure $3.3.5$).

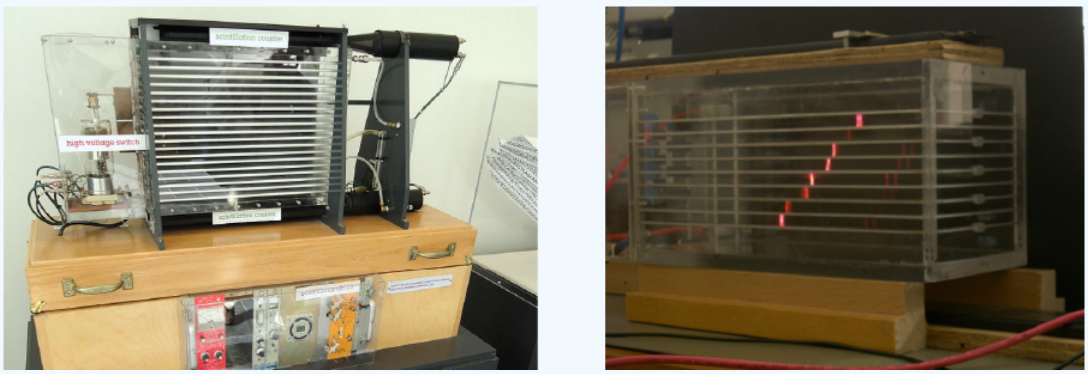


Figure $3.3.5$: A spark chamber is used to trace the paths of high-energy particles. Ionization created by the particles as they pass through the gas between the plates allows a spark to jump. The sparks are perpendicular to the plates, following electric field lines between them. The potential difference between adjacent plates is not high enough to cause sparks without the ionization produced by particles from accelerator experiments (or cosmic rays). This form of detector is now archaic and no longer in use except for demonstration purposes. (credit b: modification of work by Jack Collins)

✓ Example $3.3.1B$: Field and Force inside an Electron Gun

An electron gun (Figure $3.3.2$) has parallel plates separated by 4.00 cm and gives electrons 25.0 keV of energy. (a) What is the electric field strength between the plates? (b) What force would this field exert on a piece of plastic with a $0.500\ \mu\mathrm{C}$ charge that gets between the plates?

Strategy

Since the voltage and plate separation are given, the electric field strength can be calculated directly from the expression $E = \frac{V_{AB}}{d}$. Once we know the electric field strength, we can find the force on a charge by using $\vec{F} = q\vec{E}$. Since the electric field is in only one direction, we can write this equation in terms of the magnitudes, $F = qE$.

Solution

a. The expression for the magnitude of the electric field between two uniform metal plates is

$$E = \frac{V_{AB}}{d}.$$

Since the electron is a single charge and is given 25.0 keV of energy, the potential difference must be 25.0 kV. Entering this value for V_{AB} and the plate separation of 0.0400 m, we obtain

$$E = \frac{25.0\,\mathrm{kV}}{0.0400\,\mathrm{m}} = 6.25 \times 10^5\ \mathrm{V/m}.$$

b. The magnitude of the force on a charge in an electric field is obtained from the equation

$$F = qE.$$

Substituting known values gives

$$F = (0.500 \times 10^{-6}\,\mathrm{C})(6.25 \times 10^5\,\mathrm{V/m}) = 0.313\,\mathrm{N}.$$

Significance Note that the units are newtons, since $1\,\mathrm{V/m} = 1\,\mathrm{N/C}$. Because the electric field is uniform between the plates, the force on the charge is the same no matter where the charge is located between the plates.

Example 3.3.4C: Calculating Potential of a Point Charge

Given a point charge $q = +2.0\,\mathrm{nC}$ at the origin, calculate the potential difference between point P_1 a distance $a = 4.0\,\mathrm{cm}$ from q, and P_2 a distance $b = 12.0\,\mathrm{cm}$ from q, where the two points have an angle of $\varphi = 24^\circ$ between them (Figure 3.3.6).

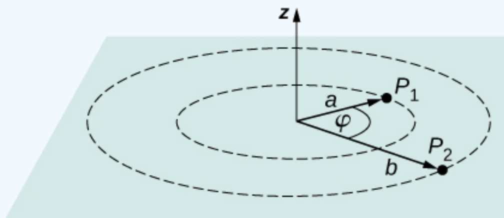


Figure 3.3.6: Find the difference in potential between P_1 and P_2.

Strategy Do this in two steps. The first step is to use $V_B - V_A = -\int_A^B \vec{E} \cdot d\vec{l}$ and let $A = a = 4.0\,\mathrm{cm}$ and $B = b = 12.0\,\mathrm{cm}$, with $d\vec{l} = d\vec{r} = \hat{r}\,dr$ and $\vec{E} = \frac{kq}{r^2}\hat{r}$. Then perform the integral. The second step is to integrate $V_B - V_A = -\int_A^B \vec{E} \cdot d\vec{l}$ around an arc of constant radius r, which means we let $d\vec{l} = r\vec{\varphi}d\varphi$ with limits $0 \leq \varphi \leq 24^\circ$, still using $\vec{E} = \frac{kq}{r^2}\hat{r}$.

Then add the two results together.

Solution For the first part, $V_B - V_A = -\int_A^B \vec{E} \cdot d\vec{l}$ for this system becomes $V_b - V_a = -\int_a^b \frac{kq}{r^2}\hat{r} \cdot \hat{r}\,dr$ which computes to

$$\Delta V = -\int_a^b \frac{kq}{r^2} dr = kq\left[\frac{1}{a} - \frac{1}{b}\right]$$

$$= (8.99 \times 10^9\,\mathrm{N\,m^2/C^2})(2.0 \times 10^{-9}\,\mathrm{C})\left[\frac{1}{0.040\,\mathrm{m}} - \frac{1}{0.12\,\mathrm{m}}\right] = 300\,\mathrm{V}.$$

For the second step, $V_B - V_A = -\int_A^B \vec{E} \cdot d\vec{l}$ becomes $\Delta V = -\int_{0^o}^{24^\circ} \frac{kq}{r^2}\hat{r} \cdot r\hat{\varphi}d\varphi$, but $\hat{r} \cdot \hat{\varphi} = 0$ and therefore $\Delta V = 0$. Adding the two parts together, we get 300 V.

Significance

We have demonstrated the use of the integral form of the potential difference to obtain a numerical result. Notice that, in this particular system, we could have also used the formula for the potential due to a point charge at the two points and simply taken the difference.

Exercise 3.3.4

From the examples, how does the energy of a lightning strike vary with the height of the clouds from the ground? Consider the cloud-ground system to be two parallel plates.

Answer

Given a fixed maximum electric field strength, the potential at which a strike occurs increases with increasing height above the ground. Hence, each electron will carry more energy. Determining if there is an effect on the total number of electrons lies in the future.

Before presenting problems involving electrostatics, we suggest a problem-solving strategy to follow for this topic.

Problem-Solving Strategy: Electrostatics

1. Examine the situation to determine if static electricity is involved; this may concern separated stationary charges, the forces among them, and the electric fields they create.
2. Identify the system of interest. This includes noting the number, locations, and types of charges involved.
3. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Determine whether the Coulomb force is to be considered directly—if so, it may be useful to draw a free-body diagram, using electric field lines.
4. Make a list of what is given or can be inferred from the problem as stated (identify the knowns). It is important to distinguish the Coulomb force F from the electric field E, for example.
5. Solve the appropriate equation for the quantity to be determined (the unknown) or draw the field lines as requested.
6. Examine the answer to see if it is reasonable: Does it make sense? Are units correct and the numbers involved reasonable?

CHAPTER OVERVIEW

4: Current and Resistance

4.1: Prelude to Current and Resistance

In this chapter, we study the electrical current through a material, where the electrical current is the rate of flow of charge. We also examine a characteristic of materials known as the resistance. Resistance is a measure of how much a material impedes the flow of charge, and it will be shown that the resistance depends on temperature. In general, a good conductor, such as copper, gold, or silver, has very low resistance. Some materials, called superconductors, have zero resistance at very low temperatures.

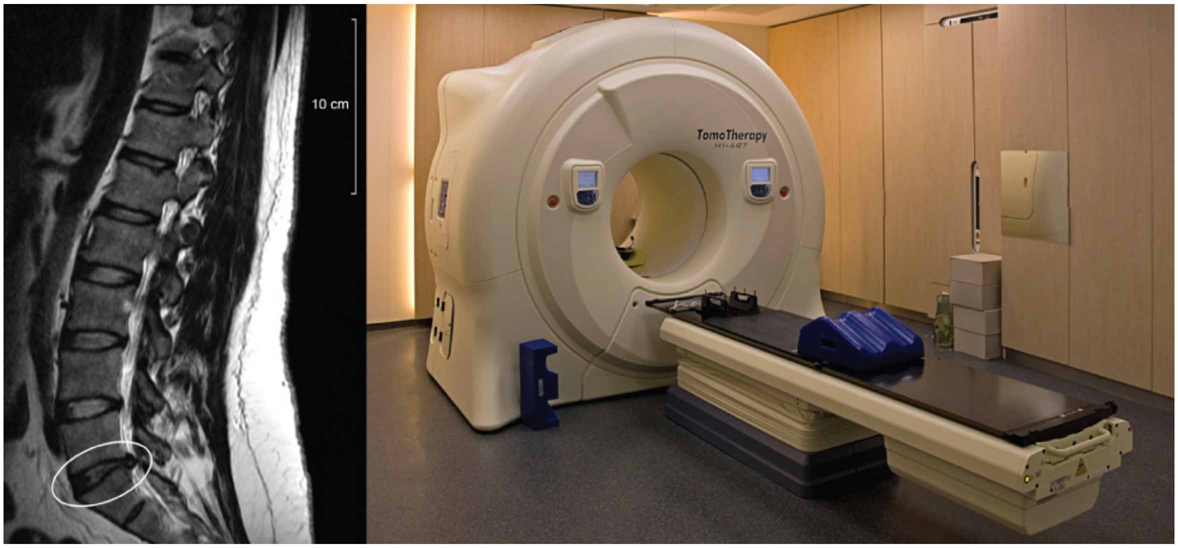


Figure 4.1.1: Magnetic resonance imaging (MRI) uses superconducting magnets and produces high-resolution images without the danger of radiation. The image on the left shows the spacing of vertebrae along a human spinal column, with the circle indicating where the vertebrae are too close due to a ruptured disc. On the right is a picture of the MRI instrument, which surrounds the patient on all sides. A large amount of electrical current is required to operate the electromagnets (credit right: modification of work by "digital cat"/Flickr).

High currents are required for the operation of electromagnets. Superconductors can be used to make electromagnets that are 10 times stronger than the strongest conventional electromagnets. These superconducting magnets are used in the construction of magnetic resonance imaging (MRI) devices that can be used to make high-resolution images of the human body. The chapter-opening picture shows an MRI image of the vertebrae of a human subject and the MRI device itself. Superconducting magnets have many other uses. For example, superconducting magnets are used in the Large Hadron Collider (LHC) to curve the path of protons in the ring.

4.2: Electrical Current

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe an electrical current
- Define the unit of electrical current
- Explain the direction of current flow

Up to now, we have considered primarily static charges. When charges did move, they were accelerated in response to an electrical field created by a voltage difference. The charges lost potential energy and gained kinetic energy as they traveled through a potential difference where the electrical field did work on the charge.

Although charges do not require a material to flow through, the majority of this chapter deals with understanding the movement of charges through a material. The rate at which the charges flow past a location—that is, the amount of charge per unit time—is known as the **electrical current**. When charges flow through a medium, the current depends on the voltage applied, the material through which the charges flow, and the state of the material. Of particular interest is the motion of charges in a conducting wire. In previous chapters, charges were accelerated due to the force provided by an electrical field, losing potential energy and gaining kinetic energy. In this chapter, we discuss the situation of the force provided by an electrical field in a conductor, where charges lose kinetic energy to the material reaching a constant velocity, known as the "**drift velocity**." This is analogous to an object falling through the atmosphere and losing kinetic energy to the air, reaching a constant terminal velocity.

If you have ever taken a course in first aid or safety, you may have heard that in the event of electric shock, it is the current, not the voltage, which is the important factor on the severity of the shock and the amount of damage to the human body. Current is measured in units called amperes; you may have noticed that circuit breakers in your home and fuses in your car are rated in amps (or amperes). But what is the ampere and what does it measure?

Defining Current and the Ampere

Electrical current is defined to be the rate at which charge flows. When there is a large current present, such as that used to run a refrigerator, a large amount of charge moves through the wire in a small amount of time. If the current is small, such as that used to operate a handheld calculator, a small amount of charge moves through the circuit over a long period of time.

Electrical Current

The average electrical current I is the rate at which charge flows,

$$I_{ave} = \frac{\Delta Q}{\Delta t}, \tag{4.2.1}$$

where ΔQ is the amount of net charge passing through a given cross-sectional area in time Δt (Figure 4.2.1). The SI unit for current is the **ampere** (A), named for the French physicist André-Marie Ampère (1775–1836). Since $I = \frac{\Delta Q}{\Delta t}$, we see that an ampere is defined as one coulomb of charge passing through a given area per second:

$$1A \equiv 1\frac{C}{s}.$$

The instantaneous electrical current, or simply the **electrical current**, is the time derivative of the charge that flows and is found by taking the limit of the average electrical current as $\Delta t \to 0$.

$$I = \lim_{\Delta t \to 0} \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}.$$

Most electrical appliances are rated in amperes (or amps) required for proper operation, as are fuses and circuit breakers.

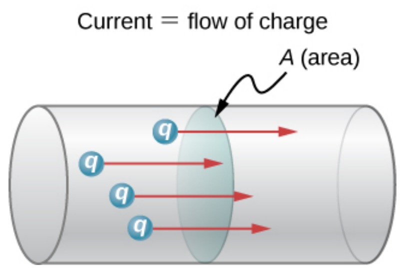


Figure 4.2.1: The rate of flow of charge is current. An ampere is the flow of one coulomb of charge through an area in one second. A current of one amp would result from 6.25×10^{18} electrons flowing through the area A each second.

culating the Average Current

The main purpose of a battery in a car or truck is to run the electric **starter motor**, which starts the engine. The operation of starting the vehicle requires a large current to be supplied by the battery. Once the engine starts, a device called an alternator takes over supplying the electric power required for running the vehicle and for charging the battery.

a. What is the average current involved when a truck battery sets in motion 720 C of charge in 4.00 s while starting an engine?
b. How long does it take 1.00 C of charge to flow from the battery?

Strategy

We can use the definition of the average current in Equation 4.2.1 to find the average current in part (a), since charge and time are given. For part (b), once we know the average current, we use Equation 4.2.1 to find the time required for 1.00 C of charge to flow from the battery.

Solution

a. Entering the given values for charge and time into the definition of current gives

$$\begin{aligned} I &= \frac{\Delta Q}{\Delta t} \\ &= \frac{720\,C}{4.00\,s} \\ &= 180\,C/s \\ &= 180\,A. \end{aligned}$$

b. Solving the relationship $I = \frac{\Delta Q}{\Delta t}$ for time Δt and entering the known values for charge and current gives

$$\begin{aligned} \Delta t &= \frac{\Delta Q}{I} \\ &= \frac{1.00\,C}{180\,C/s} \\ &= 5.56 \times 10^{-3}\,s \\ &= 5.56\,ms. \end{aligned}$$

Significance

a. This large value for current illustrates the fact that a large charge is moved in a small amount of time. The currents in these "starter motors" are fairly large to overcome the inertia of the engine.
b. A high current requires a short time to supply a large amount of charge. This large current is needed to supply the large amount of energy needed to start the engine.

Calculating Instantaneous Currents

Consider a charge moving through a cross-section of a wire where the charge is modeled as $Q(t) = Q_M(1 - e^{-t/\tau})$. Here, Q_M is the charge after a long period of time, as time approaches infinity, with units of coulombs, and τ is a time constant with units of seconds (Figure $4.2.2$). What is the current through the wire?

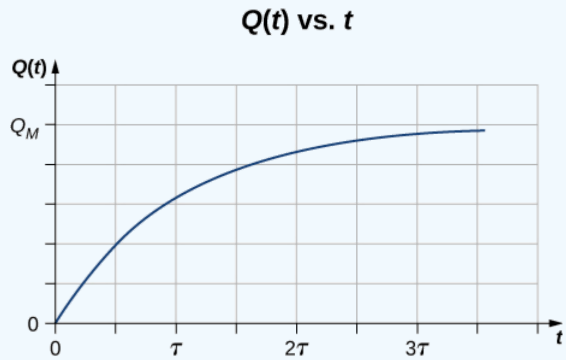


Figure $4.2.2$: A graph of the charge moving through a cross-section of a wire over time.

Strategy

The current through the cross-section can be found from $I = \frac{dQ}{dt}$. Notice from the figure that the charge increases to Q_M and the derivative decreases, approaching zero, as time increases (Figure $4.2.2$).

Solution

The derivative can be found using $\frac{d}{dx}e_u = e^u \frac{du}{dx}$.

$$\begin{aligned} I &= \frac{dQ}{dt} \\ &= \frac{d}{dt}\left[Q_M\left(1 - e^{-t/\tau}\right)\right] \\ &= \frac{Q_M}{\tau}e^{-t/\tau}. \end{aligned}$$

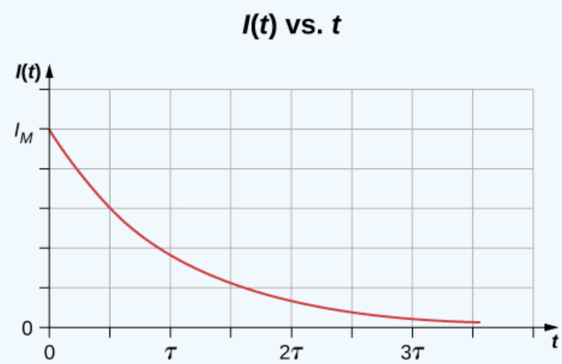


Figure $4.2.3$: A graph of the current flowing through the wire over time.

Significance

The current through the wire in question decreases exponentially, as shown in Figure $4.2.3$. In later chapters, it will be shown that a time-dependent current appears when a capacitor charges or discharges through a resistor. Recall that a capacitor is a device that stores charge. You will learn about the resistor in Model of Conduction in Metals.

? Exercise $4.2.1A$

Handheld calculators often use small solar cells to supply the energy required to complete the calculations needed to complete your next physics exam. The current needed to run your calculator can be as small as 0.30 mA. How long would it take for 1.00 C of charge to flow from the solar cells? Can solar cells be used, instead of batteries, to start traditional internal combustion engines presently used in most cars and trucks?

Answer

The time for 1.00 C of charge to flow would be

$$\Delta t = \frac{\Delta Q}{I} = \frac{1.00\,C}{0.300 \times 10^{-3} C/s} = 3.33 \times 10^3\ s.$$

This is slightly less than an hour. This is quite different from the 5.55 ms for the truck battery. The calculator takes a very small amount of energy to operate, unlike the truck's starter motor. There are several reasons that vehicles use batteries and not solar cells. Aside from the obvious fact that a light source to run the solar cells for a car or truck is not always available, the large amount of current needed to start the engine cannot easily be supplied by present-day solar cells. Solar cells can possibly be used to charge the batteries. Charging the battery requires a small amount of energy when compared to the energy required to run the engine and the other accessories such as the heater and air conditioner. Present day solar-powered cars are powered by solar panels, which may power an electric motor, instead of an internal combustion engine.

? Exercise $4.2.1B$

Circuit breakers in a home are rated in amperes, normally in a range from 10 amps to 30 amps, and are used to protect the residents from harm and their appliances from damage due to large currents. A single 15-amp circuit breaker may be used to protect several outlets in the living room, whereas a single 20-amp circuit breaker may be used to protect the refrigerator in the kitchen. What can you deduce from this about current used by the various appliances?

Answer

The total current needed by all the appliances in the living room (a few lamps, a television, and your laptop) draw less current and require less power than the refrigerator.

Current in a Circuit

In the previous paragraphs, we defined the current as the charge that flows through a cross-sectional area per unit time. In order for charge to flow through an appliance, such as the headlight shown in Figure $4.2.4$, there must be a complete path (or **circuit**) from the positive terminal to the negative terminal. Consider a simple circuit of a car battery, a switch, a headlight lamp, and wires that provide a current path between the components. In order for the lamp to light, there must be a complete path for current flow. In other words, a charge must be able to leave the positive terminal of the battery, travel through the component, and back to the negative terminal of the battery. The switch is there to control the circuit. Part (a) of the figure shows the simple circuit of a car battery, a switch, a conducting path, and a headlight lamp. Also shown is the **schematic** of the circuit [part (b)]. A schematic is a graphical representation of a circuit and is very useful in visualizing the main features of a circuit. Schematics use standardized symbols to represent the components in a circuits and solid lines to represent the wires connecting the components. The battery is shown as a series of long and short lines, representing the historic voltaic pile. The lamp is shown as a circle with a loop inside, representing the filament of an incandescent bulb. The switch is shown as two points with a conducting bar to connect the two points and the wires connecting the components are shown as solid lines. The schematic in part (c) shows the direction of current flow when the switch is closed.

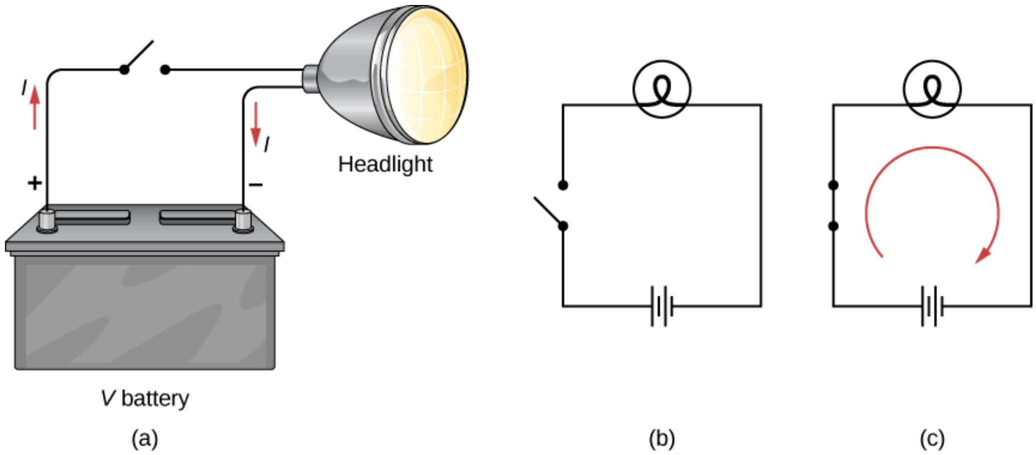


Figure 4.2.4: (a) A simple electric circuit of a headlight (lamp), a battery, and a switch. When the switch is closed, an uninterrupted path for current to flow through is supplied by conducting wires connecting a load to the terminals of a battery. (b) In this schematic, the battery is represented by parallel lines, which resemble plates in the original design of a battery. The longer lines indicate the positive terminal. The conducting wires are shown as solid lines. The switch is shown, in the open position, as two terminals with a line representing a conducting bar that can make contact between the two terminals. The lamp is represented by a circle encompassing a filament, as would be seen in an incandescent light bulb. (c) When the switch is closed, the circuit is complete and current flows from the positive terminal to the negative terminal of the battery.

When the switch is closed in Figure 4.2.4c, there is a complete path for charges to flow, from the positive terminal of the battery, through the switch, then through the headlight and back to the negative terminal of the battery. Note that the direction of current flow is from positive to negative. The direction of conventional current is always represented in the direction that positive charge would flow, from the positive terminal to the negative terminal.

The conventional current flows from the positive terminal to the negative terminal, but depending on the actual situation, positive charges, negative charges, or both may move. In metal wires, for example, current is carried by electrons—that is, negative charges move. In ionic solutions, such as salt water, both positive and negative charges move. This is also true in nerve cells. A Van de Graaff generator, used for nuclear research, can produce a current of pure positive charges, such as protons. In the Tevatron Accelerator at Fermilab, before it was shut down in 2011, beams of protons and antiprotons traveling in opposite directions were collided. The protons are positive and therefore their current is in the same direction as they travel. The antiprotons are negativity charged and thus their current is in the opposite direction that the actual particles travel.

A closer look at the current flowing through a wire is shown in Figure 4.2.5. The figure illustrates the movement of charged particles that compose a current. The fact that conventional current is taken to be in the direction that positive charge would flow can be traced back to American scientist and statesman Benjamin Franklin in the 1700s. Having no knowledge of the particles that make up the atom (namely the proton, electron, and neutron), Franklin believed that electrical current flowed from a material that had more of an “electrical fluid” and to a material that had less of this “electrical fluid.” He coined the term **positive** for the material that had more of this electrical fluid and **negative** for the material that lacked the electrical fluid. He surmised that current would flow from the material with more electrical fluid—the positive material—to the negative material, which has less electrical fluid. Franklin called this direction of current a positive current flow. This was pretty advanced thinking for a man who knew nothing about the atom.

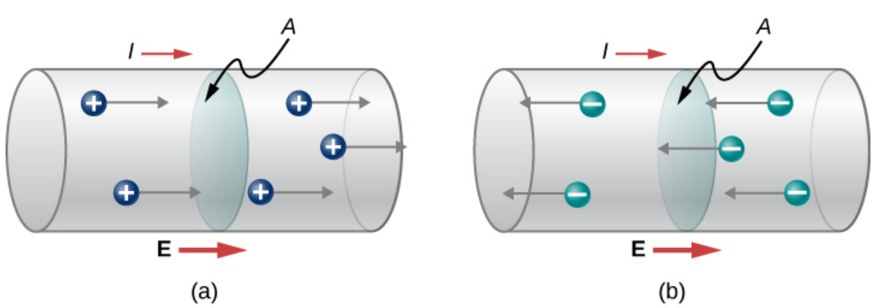


Figure 4.2.5: Current **I** is the rate at which charge moves through an area **A**, such as the cross-section of a wire. Conventional current is defined to move in the direction of the electrical field. (a) Positive charges move in the direction of the electrical field, which is the same direction as conventional current. (b) Negative charges move in the direction opposite to the electrical field. Conventional current is in the direction opposite to the movement of negative charge. The flow of electrons is sometimes referred to as electronic flow.

We now know that a material is positive if it has a greater number of protons than electrons, and it is negative if it has a greater number of electrons than protons. In a conducting metal, the current flow is due primarily to electrons flowing from the negative material to the positive material, but for historical reasons, we consider the positive current flow and the current is shown to flow from the positive terminal of the battery to the negative terminal.

It is important to realize that an electrical field is present in conductors and is responsible for producing the current (Figure $4.2.5$). In previous chapters, we considered the static electrical case, where charges in a conductor quickly redistribute themselves on the surface of the conductor in order to cancel out the external electrical field and restore equilibrium. In the case of an electrical circuit, the charges are prevented from ever reaching equilibrium by an external source of electric potential, such as a battery. The energy needed to move the charge is supplied by the electric potential from the battery.

Although the electrical field is responsible for the motion of the charges in the conductor, the work done on the charges by the electrical field does not increase the kinetic energy of the charges. We will show that the electrical field is responsible for keeping the electric charges moving at a "drift velocity."

4.3: Model of Conduction in Metals

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Define the drift velocity of charges moving through a metal
- Define the vector current density
- Describe the operation of an incandescent lamp

When electrons move through a conducting wire, they do not move at a constant velocity, that is, the electrons do not move in a straight line at a constant speed. Rather, they interact with and collide with atoms and other free electrons in the conductor. Thus, the electrons move in a zig-zag fashion and drift through the wire. We should also note that even though it is convenient to discuss the direction of current, current is a scalar quantity. When discussing the velocity of charges in a current, it is more appropriate to discuss the current density. We will come back to this idea at the end of this section.

Drift Velocity

Electrical signals move very rapidly. Telephone conversations carried by currents in wires cover large distances without noticeable delays. Lights come on as soon as a light switch is moved to the 'on' position. Most electrical signals carried by currents travel at speeds on the order of $10^8 m/s$ a significant fraction of the speed of light. Interestingly, the individual charges that make up the current move much slower on average, typically drifting at speeds on the order of $10^{-4} m/s$. How do we reconcile these two speeds, and what does it tell us about standard conductors?

The high speed of electrical signals results from the fact that the force between charges acts rapidly at a distance. Thus, when a free charge is forced into a wire, as in Figure $4.3.1$, the incoming charge pushes other charges ahead of it due to the repulsive force between like charges. These moving charges push on charges farther down the line. The density of charge in a system cannot easily be increased, so the signal is passed on rapidly. The resulting electrical shock wave moves through the system at nearly the speed of light. To be precise, this fast-moving signal, or shock wave, is a rapidly propagating change in the electrical field.

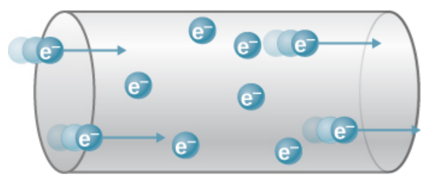


Figure $4.3.1$: When charged particles are forced into this volume of a conductor, an equal number are quickly forced to leave. The repulsion between like charges makes it difficult to increase the number of charges in a volume. Thus, as one charge enters, another leaves almost immediately, carrying the signal rapidly forward.

Good conductors have large numbers of free charges. In metals, the free charges are free electrons. (In fact, good electrical conductors are often good heat conductors too, because large numbers of free electrons can transport thermal energy as well as carry electrical current.) Figure $4.3.2$ shows how free electrons move through an ordinary conductor. The distance that an individual electron can move between collisions with atoms or other electrons is quite small. The electron paths thus appear nearly random, like the motion of atoms in a gas. But there is an electrical field in the conductor that causes the electrons to drift in the direction shown (opposite to the field, since they are negative). The **drift velocity** $\vec{v}_d$ is the average velocity of the free charges. Drift velocity is quite small, since there are so many free charges. If we have an estimate of the density of free electrons in a conductor, we can calculate the drift velocity for a given current. The larger the density, the lower the velocity required for a given current.

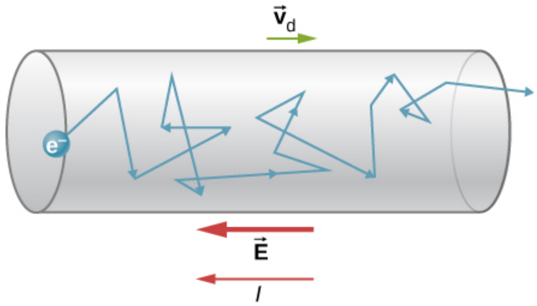


Figure 4.3.2: Free electrons moving in a conductor make many collisions with other electrons and other particles. A typical path of one electron is shown. The average velocity of the free charges is called the drift velocity $\vec{v}_d$ and for electrons, it is in the direction opposite to the electrical field. The collisions normally transfer energy to the conductor, requiring a constant supply of energy to maintain a steady current.

Free-electron collisions transfer energy to the atoms of the conductor. The electrical field does work in moving the electrons through a distance, but that work does not increase the kinetic energy (nor speed) of the electrons. The work is transferred to the conductor's atoms, often increasing temperature. Thus, a continuous power input is required to keep a current flowing. (An exception is superconductors, for reasons we shall explore in a later chapter. Superconductors can have a steady current without a continual supply of energy—a great energy savings.) For a conductor that is not a superconductor, the supply of energy can be useful, as in an incandescent light bulb filament (Figure 4.3.3). The supply of energy is necessary to increase the temperature of the tungsten filament, so that the filament glows.

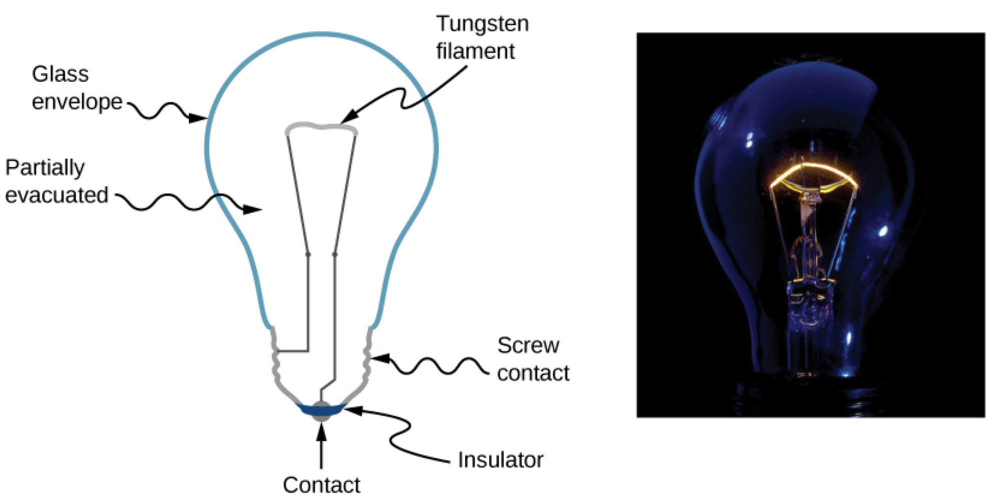


Figure 4.3.3: The incandescent lamp is a simple design. A tungsten filament is placed in a partially evacuated glass envelope. One end of the filament is attached to the screw base, which is made out of a conducting material. The second end of the filament is attached to a second contact in the base of the bulb. The two contacts are separated by an insulating material. Current flows through the filament, and the temperature of the filament becomes large enough to cause the filament to glow and produce light. However, these bulbs are not very energy efficient, as evident from the heat coming from the bulb. In the year 2012, the United States, along with many other countries, began to phase out incandescent lamps in favor of more energy-efficient lamps, such as light-emitting diode (LED) lamps and compact fluorescent lamps (CFL) (credit right: modification of work by Serge Saint).

We can obtain an expression for the relationship between current and drift velocity by considering the number of free charges in a segment of wire, as illustrated in Figure 4.3.4. The number of free charges per unit volume, or the number density of free charges, is given the symbol n where

$$n = \frac{\text{number of charges}}{\text{volume}}.$$

The value of n depends on the material. The shaded segment has a volume $Av_d\,dt$, so that the number of free charges in the volume is $nAv_d\,dt$. The charge dQ in this segment is thus $qnAv_d\,dt$, where q is the amount of charge on each carrier. (The magnitude of the charge of electrons is $q = 1.60 \times 10^{-19}\,C$.) Current is charge moved per unit time; thus, if all the original charges move out of this segment in time **dt**, the current is

$$I = \frac{dQ}{dt} = qnAv_d.$$

Rearranging terms gives

$$v_d = \frac{I}{nqA}$$

where

- v_d is the drift velocity,
- $\boldsymbol{n}$ is the free charge density,
- $\boldsymbol{A}$ is the cross-sectional area of the wire, and
- I is the current through the wire.

The carriers of the current each have charge **q** and move with a drift velocity of magnitude v_d.

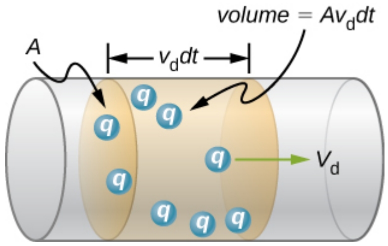


Figure 4.3.4: All the charges in the shaded volume of this wire move out in a time **dt**, having a drift velocity of magnitude v_d.

Note that simple drift velocity is not the entire story. The speed of an electron is sometimes much greater than its drift velocity. In addition, not all of the electrons in a conductor can move freely, and those that do move might move somewhat faster or slower than the drift velocity. So what do we mean by free electrons?

Atoms in a metallic conductor are packed in the form of a lattice structure. Some electrons are far enough away from the atomic nuclei that they do not experience the attraction of the nuclei as strongly as the inner electrons do. These are the free electrons. They are not bound to a single atom but can instead move freely among the atoms in a "sea" of electrons. When an electrical field is applied, these free electrons respond by accelerating. As they move, they collide with the atoms in the lattice and with other electrons, generating thermal energy, and the conductor gets warmer. In an insulator, the organization of the atoms and the structure do not allow for such free electrons.

As you know, electric power is usually supplied to equipment and appliances through round wires made of a conducting material (copper, aluminum, silver, or gold) that are stranded or solid. The diameter of the wire determines the current-carrying capacity—the larger the diameter, the greater the current-carrying capacity. Even though the current-carrying capacity is determined by the diameter, wire is not normally characterized by the diameter directly. Instead, wire is commonly sold in a unit known as "gauge." Wires are manufactured by passing the material through circular forms called "drawing dies." In order to make thinner wires, manufacturers draw the wires through multiple dies of successively thinner diameter. Historically, the gauge of the wire was related to the number of drawing processes required to manufacture the wire. For this reason, the larger the gauge, the smaller the diameter. In the United States, the American Wire Gauge (AWG) was developed to standardize the system. Household wiring commonly consists of 10-gauge (2.588-mm diameter) to 14-gauge (1.628-mm diameter) wire. A device used to measure the gauge of wire is shown in Figure 4.3.5.

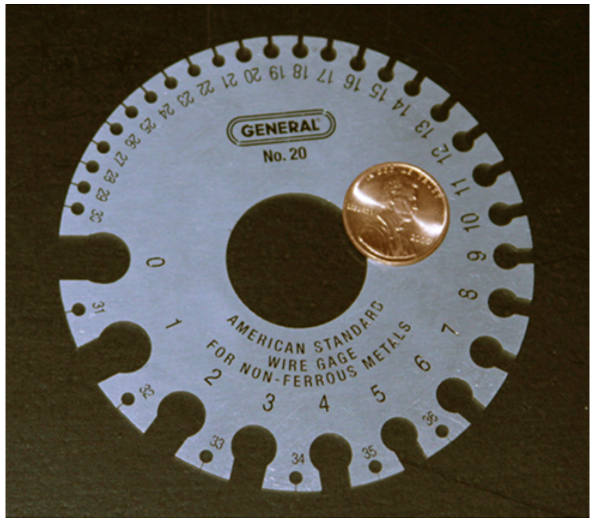


Figure 4.3.5: A device for measuring the gauge of electrical wire. As you can see, higher gauge numbers indicate thinner wires.

Example 4.3.1: Calculating Drift Velocity in a Common Wire

Calculate the drift velocity of electrons in a copper wire with a diameter of 2.053 mm (12-gauge) carrying a 20.0-A current, given that there is one free electron per copper atom. (Household wiring often contains 12-gauge copper wire, and the maximum current allowed in such wire is usually 20.0 A.) The density of copper is $8.80 \times 10^3 kg/m^3$ and the atomic mass of copper is 63.54 g/mol.

Strategy

We can calculate the drift velocity using the equation $I = nqAv_d$. The current is $I = 20.00\,A$ and $q = 1.60 \times 10{-}19C$ is the charge of an electron. We can calculate the area of a cross-section of the wire using the formula $A = \pi r^2$, where r is one-half the diameter. The given diameter is 2.053 mm, so **r** is 1.0265 mm. We are given the density of copper, $8.80 \times 10^3 kg/m^3$, and the atomic mass of copper is $63.54\,g/mol$. We can use these two quantities along with Avogadro's number, $6.02 \times 10^{23}\,atoms/mol$, to determine **n**, the number of free electrons per cubic meter.

Solution

First, we calculate the density of free electrons in copper. There is one free electron per copper atom. Therefore, the number of free electrons is the same as the number of copper atoms per m^3. We can now find **n** as follows:

$$\begin{aligned} n &= \frac{1\,e^-}{atom} \times \frac{6.02 \times 10^{23} atoms}{mol} \times \frac{1\,mol}{63.54\,g} \times \frac{1000\,g}{kg} \times \frac{8.80 \times 10^3 kg}{1\,m^3} \\ &= 8.34 \times 10^{28} e^-/m^3. \end{aligned}$$

The cross-sectional area of the wire is

$$\begin{aligned} A &= \pi r^2 \\ &= \pi \left(\frac{2.05 \times 10^{-3} m}{2}\right)^2 \\ &= 3.30 \times 10^{-6} m^2. \end{aligned}$$

Rearranging $I = nqAv_d$ to isolate drift velocity gives

$$\begin{aligned} v_d &= \frac{I}{nqA} \\ &= \frac{20.00\,A}{(8.34 \times 10^{28}/m^3)(-1.60 \times 10^{-19} C)(3.30 \times 10^{-6} m^2)} \\ &= -4.54 \times 10^{-4} m/s. \end{aligned}$$

Significance

The minus sign indicates that the negative charges are moving in the direction opposite to conventional current. The small value for drift velocity (on the order of $10^{-4} m/s$ confirms that the signal moves on the order of 10^{12} times faster (about $10^8 m/s$ than the charges that carry it.

? Exercise 4.3.1

In Example 4.3.1, the drift velocity was calculated for a 2.053-mm diameter (12-gauge) copper wire carrying a 20-amp current. Would the drift velocity change for a 1.628-mm diameter (14-gauge) wire carrying the same 20-amp current?

Answer

The diameter of the 14-gauge wire is smaller than the diameter of the 12-gauge wire. Since the drift velocity is inversely proportional to the cross-sectional area, the drift velocity in the 14-gauge wire is larger than the drift velocity in the 12-gauge wire carrying the same current. The number of electrons per cubic meter will remain constant.

Current Density

Although it is often convenient to attach a negative or positive sign to indicate the overall direction of motion of the charges, current is a scalar quantity, $I = \frac{dQ}{dt}$. It is often necessary to discuss the details of the motion of the charge, instead of discussing the overall motion of the charges. In such cases, it is necessary to discuss the current density, $\vec{J}$, a vector quantity. The **current density** is the flow of charge through an infinitesimal area, divided by the area. The current density must take into account the local magnitude and direction of the charge flow, which varies from point to point. The unit of current density is ampere per meter squared, and the direction is defined as the direction of net flow of positive charges through the area.

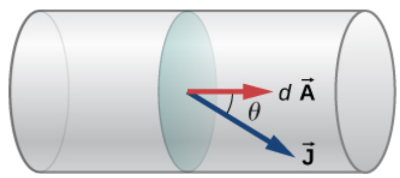


Figure 4.3.6: The current density $\vec{J}$ is defined as the current passing through an infinitesimal cross-sectional area divided by the area. The direction of the current density is the direction of the net flow of positive charges and the magnitude is equal to the current divided by the infinitesimal area.

The relationship between the current and the current density can be seen in Figure 4.3.6. The differential current flow through the area $d\vec{A}$ is found as

$$dI = \vec{J} \cdot d\vec{A} = JdA \cos \theta,$$

where θ is the angle between the area and the current density. The total current passing through area $d\vec{A}$ can be found by integrating over the area,

$$I = \iint_{area} \vec{J} \cdot d\vec{A}.$$

Consider the magnitude of the current density, which is the current divided by the area:

$$J = \frac{I}{A} = \frac{n|q|Av_d}{A} = n|q|v_d.$$

Thus, the current density is $\vec{J} = nq\vec{v}_d$. If **q** is positive, $\vec{v}_d$ is in the same direction as the electrical field $\vec{E}$. If **q** is negative, $\vec{v}_d$ is in the opposite direction of $\vec{E}$. Either way, the direction of the current density $\vec{J}$ is in the direction of the electrical field $\vec{E}$.

Example 4.3.2: Calculating the Current Density in a Wire

The current supplied to a lamp with a 100-W light bulb is 0.87 amps. The lamp is wired using a copper wire with diameter 2.588 mm (10-gauge). Find the magnitude of the current density.

Strategy

The current density is the current moving through an infinitesimal cross-sectional area divided by the area. We can calculate the magnitude of the current density using $J = \frac{I}{A}$. The current is given as 0.87 A. The cross-sectional area can be calculated to be $A = 5.26\, mm^2$.

Solution

Calculate the current density using the given current $I = 0.87\, A$ and the area, found to be $A = 5.26\, mm^2$.

$$J = \frac{I}{A} = \frac{0.87\, A}{5.26 \times 10^{-6} m^2} = 1.65 \times 10^5 \frac{A}{m^2}.$$

Significance

The current density in a conducting wire depends on the current through the conducting wire and the cross-sectional area of the wire. For a given current, as the diameter of the wire increases, the charge density decreases.

Exercise 4.3.2

The current density is proportional to the current and inversely proportional to the area. If the current density in a conducting wire increases, what would happen to the drift velocity of the charges in the wire?

Answer

The current density in a conducting wire increases due to an increase in current. The drift velocity is inversely proportional to the current $\left(v_d = \frac{nqA}{I}\right)$, so the drift velocity would decrease.

What is the significance of the current density? The current density is proportional to the current, and the current is the number of charges that pass through a cross-sectional area per second. The charges move through the conductor, accelerated by the electric force provided by the electrical field. The electrical field is created when a voltage is applied across the conductor. In Ohm's Law, we will use this relationship between the current density and the electrical field to examine the relationship between the current through a conductor and the voltage applied.

4.4: Resistivity and Resistance

Learning Objectives

By the end of this section, you will be able to:

- Differentiate between resistance and resistivity
- Define the term conductivity
- Describe the electrical component known as a resistor
- State the relationship between resistance of a resistor and its length, cross-sectional area, and resistivity
- State the relationship between resistivity and temperature

What drives current? We can think of various devices—such as batteries, generators, wall outlets, and so on—that are necessary to maintain a current. All such devices create a potential difference and are referred to as voltage sources. When a voltage source is connected to a conductor, it applies a potential difference **V** that creates an electrical field. The electrical field, in turn, exerts force on free charges, causing current. The amount of current depends not only on the magnitude of the voltage, but also on the characteristics of the material that the current is flowing through. The material can resist the flow of the charges, and the measure of how much a material resists the flow of charges is known as the **resistivity**. This resistivity is crudely analogous to the friction between two materials that resists motion.

Resistivity

When a voltage is applied to a conductor, an electrical field $\vec{E}$ is created, and charges in the conductor feel a force due to the electrical field. The current density $\vec{J}$ that results depends on the electrical field and the properties of the material. This dependence can be very complex. In some materials, including metals at a given temperature, the current density is approximately proportional to the electrical field. In these cases, the current density can be modeled as

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

where σ is the **electrical conductivity**. The electrical conductivity is analogous to thermal conductivity and is a measure of a material's ability to conduct or transmit electricity. Conductors have a higher electrical conductivity than insulators. Since the electrical conductivity is $\sigma=J/E$, the units are

$$\sigma=\frac{|J|}{|E|}=\frac{A/m^2}{V/m}=\frac{A}{V\cdot m}.$$

Here, we define a unit named the **ohm** with the Greek symbol uppercase omega, Ω. The unit is named after Georg Simon Ohm, whom we will discuss later in this chapter. The Ω is used to avoid confusion with the number 0. One ohm equals one volt per amp: $1\,\Omega=1\,V/A$. The units of electrical conductivity are therefore $(\Omega\cdot m)^{-1}$.

Conductivity is an intrinsic property of a material. Another intrinsic property of a material is the resistivity, or electrical **resistivity**. The resistivity of a material is a measure of how strongly a material opposes the flow of electrical current. The symbol for resistivity is the lowercase Greek letter rho, ρ, and resistivity is the reciprocal of electrical conductivity:

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

The unit of resistivity in SI units is the ohm-meter $(\Omega\cdot m$. We can define the resistivity in terms of the electrical field and the current density.

$$\rho=\frac{E}{J}.$$

The greater the resistivity, the larger the field needed to produce a given current density. The lower the resistivity, the larger the current density produced by a given electrical field. Good conductors have a high conductivity and low resistivity. Good insulators have a low conductivity and a high resistivity. Table $4.4.1$ lists resistivity and conductivity values for various materials.

Table $4.4.1$: Resistivities and Conductivities of Various Materials at 20 °C[1] Values depend strongly on amounts and types of impurities.

Material	Conductivity, σ $(\Omega \cdot m)^{-1}$	Resistivity, ρ $(\Omega \cdot m)$	Temperature Coefficient α $(^oC)^{-1}$
Conductors			
Silver	6.29×10^7	1.59×10^{-8}	0.0038
Copper	5.95×10^7	1.68×10^{-8}	0.0039
Gold	4.10×10^7	2.44×10^{-8}	0.0034
Aluminum	3.77×10^7	2.65×10^{-8}	0.0039
Tungsten	1.79×10^7	5.60×10^{-8}	0.0045
Iron	1.03×10^7	9.71×10^{-8}	0.0065
Platinum	0.94×10^7	10.60×10^{-8}	0.0039
Steel	0.50×10^7	20.00×10^{-8}	
Lead	0.45×10^7	22.00×10^{-8}	
Manganin (Cu, Mn. Ni alloy)	0.21×10^7	48.20×10^{-8}	0.000002
Constantan (Cu, Ni alloy)	0.20×10^7	49.00×10^{-8}	0.00003
Mercury	0.10×10^7	98.00×10^{-8}	0.0009
Nichrome (Ni, Fe, Cr alloy)	0.10×10^7	100.00×10^{-8}	0.0004
Semiconductors [1]			
Carbon (pure)	2.86×10^4	3.50×10^{-5}	-0.0005
Carbon	$(2.86 - 1.67) \times 10^{-6}$	$(3.5 - 60) \times 10^{-5}$	-0.0005
Germanium (pure)		600×10^{-3}	-0.048
Germanium		$(1 - 600) \times 10^{-3}$	-0.050
Silicon (pure)		2300	-0.075
Silicon		0.1 - 2300	-0.07
Insulators			
Amber	2.00×10^{-15}	5×10^{14}	
Glass	$10^{-9} - 19^{-14}$	$10^9 - 10^{14}$	
Lucite	$< 10^{-13}$	$> 10^{13}$	
Mica	$10^{-11} - 10^{-15}$	$10^{11} - 10^{15}$	
Quartz (fused)	1.33×10^{-18}	75×10^{16}	
Rubber (hard)	$10^{-13} - 10^{-16}$	$10^{13} - 10^{16}$	
Sulfur	10^{-15}	10^{15}	
Teflon^{TM}	$< 10^{-13}$	$> 10^{13}$	
Wood	$10^{-8} - 10^{-11}$	$10^8 - 10^{11}$	

The materials listed in the table are separated into categories of conductors, semiconductors, and insulators, based on broad groupings of resistivity. Conductors have the smallest resistivity, and insulators have the largest; semiconductors have intermediate resistivity. Conductors have varying but large, free charge densities, whereas most charges in insulators are bound to atoms and are not free to move. Semiconductors are intermediate, having far fewer free charges than conductors, but having properties that make the number of free charges depend strongly on the type and amount of impurities in the semiconductor. These unique properties of semiconductors are put to use in modern electronics, as we will explore in later chapters.

Example 4.4.1: Current Density, Resistance, and Electrical field for a Current-Carrying Wire

Calculate the current density, resistance, and electrical field of a 5-m length of copper wire with a diameter of 2.053 mm (12-gauge) carrying a current of $I - 10\,mA$.

Strategy

We can calculate the current density by first finding the cross-sectional area of the wire, which is $A = 3.31\,mm^2$, and the definition of current density $J = \frac{I}{A}$. The resistance can be found using the length of the wire $L = 5.00\,m$, the area, and the resistivity of copper $\rho = 1.68 \times 10^{-8}\,\Omega \cdot m$, where $R = \rho\frac{L}{A}$. The resistivity and current density can be used to find the electrical field.

Solution

First, we calculate the current density:

$$\begin{aligned} J &= \frac{I}{A} \\ &= \frac{10 \times 10^{-3}\,A}{3.31 \times 10^{-6}\,m^2} \\ &= 3.02 \times 10^{3}\,\frac{A}{m^2}. \end{aligned}$$

The resistance of the wire is

$$\begin{aligned} R &= \rho\frac{L}{A} \\ &= (1.68 \times 10^{-8}\,\Omega \cdot m)\frac{5.00\,m}{3.31 \times 10^{-6}\,m^2} \\ &= 0.025\,\Omega. \end{aligned}$$

Finally, we can find the electrical field:

$$\begin{aligned} E &= \rho J \\ &= 1.68 \times 10^{-8}\,\Omega \cdot m\left(3.02 \times 10^{3}\,\frac{A}{m^2}\right) \\ &= 5.07 \times 10^{-5}\,\frac{V}{m}. \end{aligned}$$

Significance

From these results, it is not surprising that copper is used for wires for carrying current because the resistance is quite small. Note that the current density and electrical field are independent of the length of the wire, but the voltage depends on the length.

Exercise 4.4.1

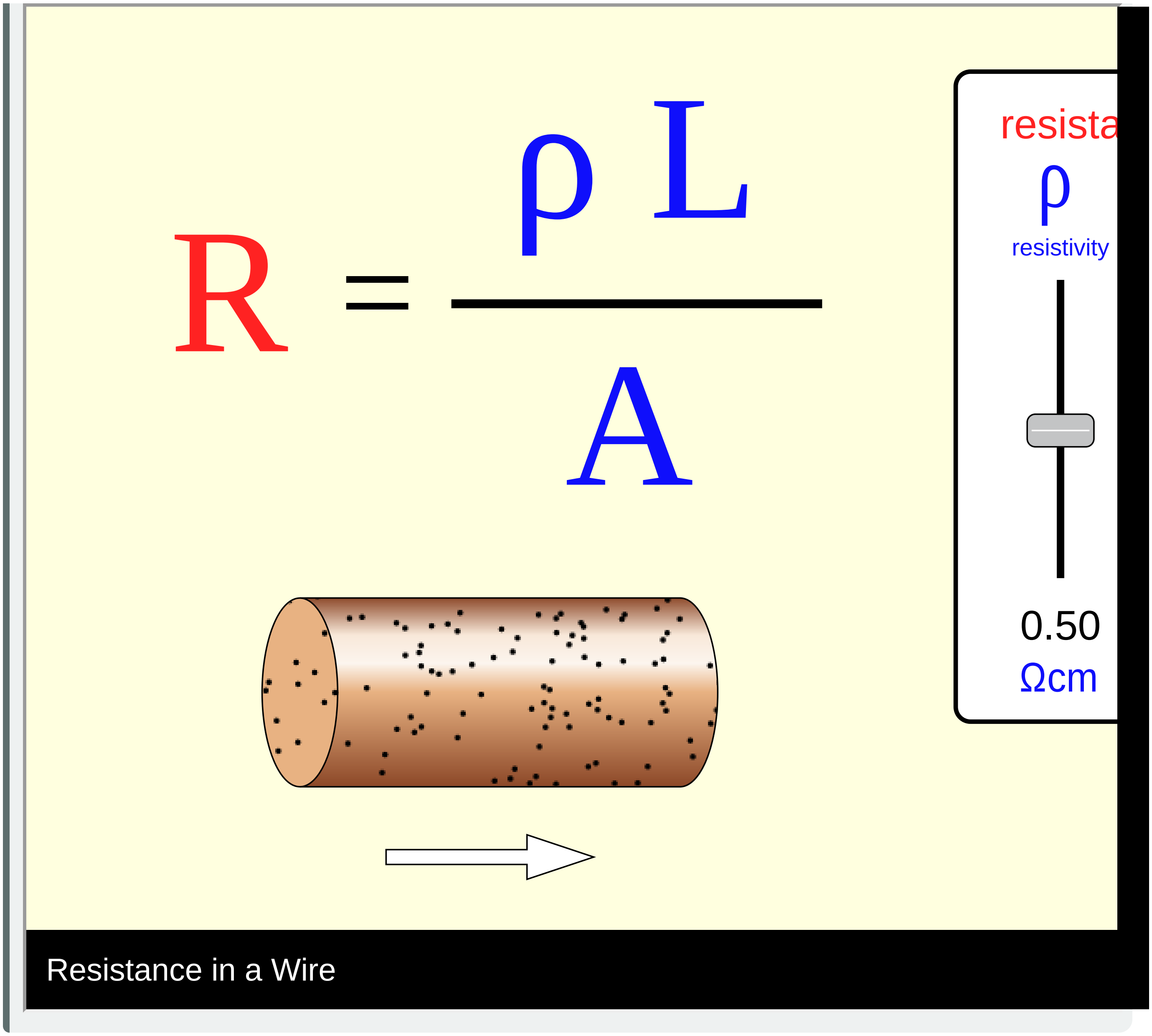
Copper wires use routinely used for extension cords and house wiring for several reasons. Copper has the highest electrical conductivity rating, and therefore the lowest resistivity rating, of all nonprecious metals. Also important is the tensile strength, where the tensile strength is a measure of the force required to pull an object to the point where it breaks. The tensile strength of a material is the maximum amount of tensile stress it can take before breaking. Copper has a high tensile strength, $2 \times 10^{8}\,\frac{N}{m^2}$. A third important characteristic is ductility. Ductility is a measure of a material's ability to be drawn into wires and a measure of the flexibility of the material, and copper has a high ductility. Summarizing, for a conductor to be a suitable candidate for making wire, there are at least three important characteristics: low resistivity, high tensile strength, and high ductility. What other materials are used for wiring and what are the advantages and disadvantages?

Answer

Silver, gold, and aluminum are all used for making wires. All four materials have a high conductivity, silver having the highest. All four can easily be drawn into wires and have a high tensile strength, though not as high as copper. The obvious disadvantage of gold and silver is the cost, but silver and gold wires are used for special applications, such as speaker wires. Gold does not oxidize, making better connections between components. Aluminum wires do have their drawbacks. Aluminum has a higher resistivity than copper, so a larger diameter is needed to match the resistance per length of copper wires, but aluminum is cheaper than copper, so this is not a major drawback. Aluminum wires do not have as high of a ductility and tensile strength as copper, but the ductility and tensile strength is within acceptable levels. There are a few concerns that must be addressed in using aluminum and care must be used when making connections. Aluminum has a higher rate of thermal expansion than copper, which can lead to loose connections and a possible fire hazard. The oxidation of aluminum does not conduct and can cause problems. Special techniques must be used when using aluminum wires and components, such as electrical outlets, must be designed to accept aluminum wires.

PhET

View this interactive simulation to see what the effects of the cross-sectional area, the length, and the resistivity of a wire are on the resistance of a conductor. Adjust the variables using slide bars and see if the resistance becomes smaller or larger.



Temperature Dependence of Resistivity

Looking back at Table 4.4.1, you will see a column labeled "Temperature Coefficient." The resistivity of some materials has a strong temperature dependence. In some materials, such as copper, the resistivity increases with increasing temperature. In fact, in most conducting metals, the resistivity increases with increasing temperature. The increasing temperature causes increased vibrations of the atoms in the lattice structure of the metals, which impede the motion of the electrons. In other materials, such as carbon, the resistivity decreases with increasing temperature. In many materials, the dependence is approximately linear and can be modeled using a linear equation:

$$\rho \approx \rho_0[1+\alpha(T-T_0)],$$

where ρ is the resistivity of the material at temperature **T**, α is the temperature coefficient of the material, and ρ_0 is the resistivity at T_0, usually taken as $T_0 = 20.00^o C$.

Note also that the temperature coefficient α is negative for the semiconductors listed in Table 4.4.1, meaning that their resistivity decreases with increasing temperature. They become better conductors at higher temperature, because increased thermal agitation increases the number of free charges available to carry current. This property of decreasing ρ with temperature is also related to the type and amount of impurities present in the semiconductors.

Resistance

We now consider the resistance of a wire or component. The resistance is a measure of how difficult it is to pass current through a wire or component. Resistance depends on the resistivity. The resistivity is a characteristic of the material used to fabricate a wire or other electrical component, whereas the resistance is a characteristic of the wire or component.

To calculate the resistance, consider a section of conducting wire with cross-sectional area **A**, length **L**, and resistivity ρ. A battery is connected across the conductor, providing a potential difference ΔV across it (Figure $4.4.1$). The potential difference produces an electrical field that is proportional to the current density, according to $\vec{E} = \rho\vec{J}$.

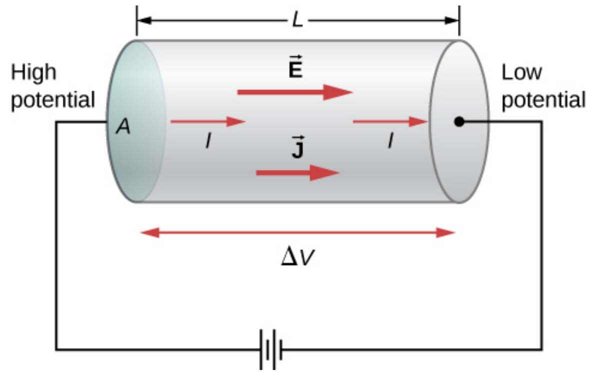


Figure $4.4.1$: A potential provided by a battery is applied to a segment of a conductor with a cross-sectional area A and a length L.

The magnitude of the electrical field across the segment of the conductor is equal to the voltage divided by the length, $E = V/L$, and the magnitude of the current density is equal to the current divided by the cross-sectional area, $J = I/A$. Using this information and recalling that the electrical field is proportional to the resistivity and the current density, we can see that the voltage is proportional to the current:

$$E = \rho J$$

$$\frac{V}{L} = \rho\frac{I}{A}$$

$$V = \left(\rho\frac{L}{A}\right) I.$$

Definition: Resistance

The ratio of the voltage to the current is defined as the **resistance** R:

$$R \equiv \frac{V}{I}.$$

The resistance of a cylindrical segment of a conductor is equal to the resistivity of the material times the length divided by the area:

$$R \equiv \frac{V}{I} = \rho\frac{L}{A}.$$

The unit of resistance is the ohm, Ω. For a given voltage, the higher the resistance, the lower the current.

Resistors

A common component in electronic circuits is the resistor. The resistor can be used to reduce current flow or provide a voltage drop. Figure $4.4.2$ shows the symbols used for a resistor in schematic diagrams of a circuit. Two commonly used standards for circuit diagrams are provided by the American National Standard Institute (ANSI, pronounced "AN-see") and the International Electrotechnical Commission (IEC). Both systems are commonly used. We use the ANSI standard in this text for its visual recognition, but we note that for larger, more complex circuits, the IEC standard may have a cleaner presentation, making it easier to read.

Figure 4.4.2: Symbols for a resistor used in circuit diagrams. (a) The ANSI symbol; (b) the IEC symbol.

Material and shape dependence of resistance

A resistor can be modeled as a cylinder with a cross-sectional area **A** and a length **L**, made of a material with a resistivity ρ (Figure 4.4.3). The resistance of the resistor is $R = \rho \frac{L}{A}$

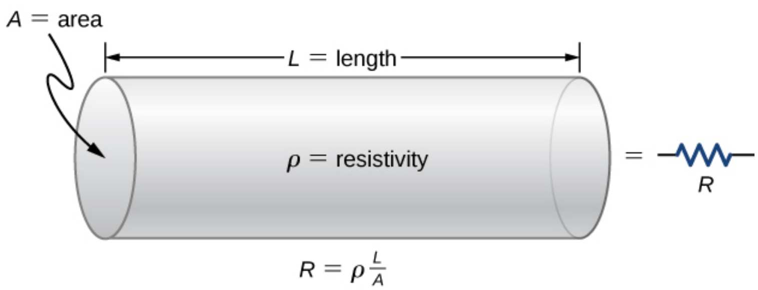


Figure 4.4.3: A model of a resistor as a uniform cylinder of length **L** and cross-sectional area **A**. Its resistance to the flow of current is analogous to the resistance posed by a pipe to fluid flow. The longer the cylinder, the greater its resistance. The larger its cross-sectional area **A**, the smaller its resistance.

The most common material used to make a resistor is carbon. A carbon track is wrapped around a ceramic core, and two copper leads are attached. A second type of resistor is the metal film resistor, which also has a ceramic core. The track is made from a metal oxide material, which has semiconductive properties similar to carbon. Again, copper leads are inserted into the ends of the resistor. The resistor is then painted and marked for identification. A resistor has four colored bands, as shown in Figure 4.4.4.

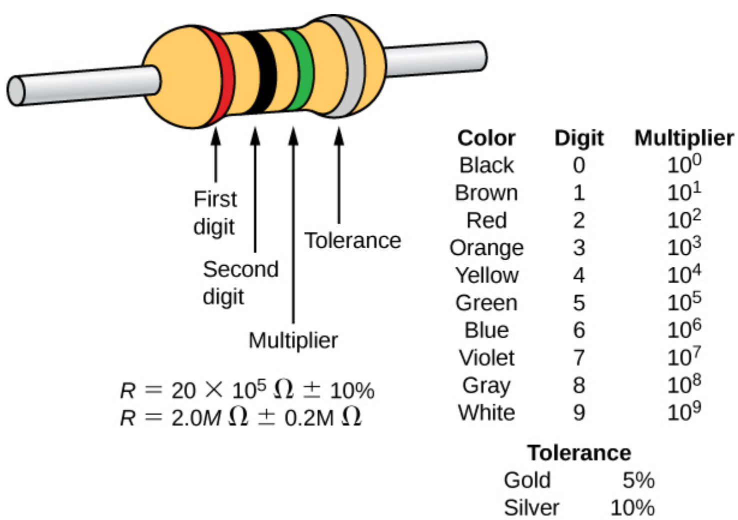


Figure 4.4.4: Many resistors resemble the figure shown above. The four bands are used to identify the resistor. The first two colored bands represent the first two digits of the resistance of the resistor. The third color is the multiplier. The fourth color represents the tolerance of the resistor. The resistor shown has a resistance of $20 \times 10^5\ \Omega \pm 10\%$

Resistances range over many orders of magnitude. Some ceramic insulators, such as those used to support power lines, have resistances of $10^{12}\ \Omega$ or more. A dry person may have a hand-to-foot resistance of $10^5\ \Omega$ whereas the resistance of the human heart is about $10^3\ \Omega$ A meter-long piece of large-diameter copper wire may have a resistance of $10^{-5}\ \Omega$, and superconductors have no resistance at all at low temperatures. As we have seen, resistance is related to the shape of an object and the material of which it is composed.

The resistance of an object also depends on temperature, since R_0 is directly proportional to ρ. For a cylinder, we know $R = \rho \frac{L}{A}$, so if **L** and **A** do not change greatly with temperature, **R** has the same temperature dependence as ρ. (Examination of the coefficients of linear expansion shows them to be about two orders of magnitude less than typical temperature coefficients of resistivity, so the effect of temperature on **L** and **A** is about two orders of magnitude less than on ρ.) Thus,

$$R = R_0(1 + \alpha \Delta T) \tag{4.4.1}$$

is the temperature dependence of the resistance of an object, where R_0 is the original resistance (usually taken to be $T = 20.00^o C$ and **R** is the resistance after a temperature change ΔT. The color code gives the resistance of the resistor at a temperature of $T = 20.00^o C$.

Numerous thermometers are based on the effect of temperature on resistance (Figure 4.4.5). One of the most common thermometers is based on the thermistor, a semiconductor crystal with a strong temperature dependence, the resistance of which is measured to obtain its temperature. The device is small, so that it quickly comes into thermal equilibrium with the part of a person it touches.

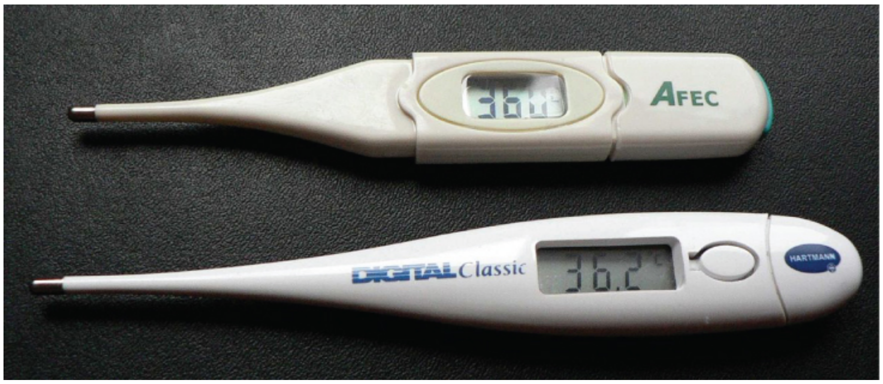


Figure 4.4.5: These familiar thermometers are based on the automated measurement of a thermistor's temperature-dependent resistance.

Example 4.4.2: Calculating Resistance

Although caution must be used in applying $\rho = \rho_0(1 + \alpha \Delta T)$ and $R = R_0(1 + \alpha \Delta T)$ for temperature changes greater than $100^o C$, for tungsten, the equations work reasonably well for very large temperature changes. A tungsten filament at $20^o C$ has a resistance of $0.350\,\Omega$. What would the resistance be if the temperature is increased to $2850^o C$?

Strategy

This is a straightforward application of Equation 4.4.1, since the original resistance of the filament is given as $R_0 = 0.350\,\Omega$ and the temperature change is $\Delta T = 2830^o C$.

Solution

The resistance of the hotter filament R is obtained by entering known values into the above equation:

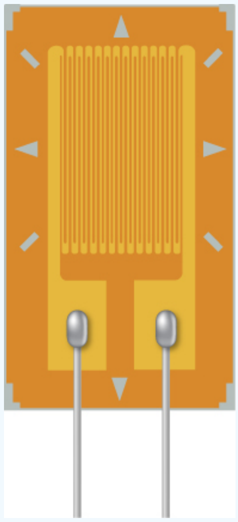
$$\begin{aligned} R &= R_0(1 + \alpha \Delta T) \\ &= (0.350\,\Omega)\left(1 + \left(\frac{4.5 \times 10^{-3}}{^oC}\right)(2830^o C)\right) \\ &= 4.8\,\Omega \end{aligned}$$

Significance

Notice that the resistance changes by more than a factor of 10 as the filament warms to the high temperature and the current through the filament depends on the resistance of the filament and the voltage applied. If the filament is used in an incandescent light bulb, the initial current through the filament when the bulb is first energized will be higher than the current after the filament reaches the operating temperature.

Exercise 4.4.2

A strain gauge is an electrical device to measure strain, as shown below. It consists of a flexible, insulating backing that supports a conduction foil pattern. The resistance of the foil changes as the backing is stretched. How does the strain gauge resistance change? Is the strain gauge affected by temperature changes?



Answer

The foil pattern stretches as the backing stretches, and the foil tracks become longer and thinner. Since the resistance is calculated as $R=\rho\frac{L}{A}$, the resistance increases as the foil tracks are stretched. When the temperature changes, so does the resistivity of the foil tracks, changing the resistance. One way to combat this is to use two strain gauges, one used as a reference and the other used to measure the strain. The two strain gauges are kept at a constant temperature

The Resistance of Coaxial Cable

Long cables can sometimes act like antennas, picking up electronic noise, which are signals from other equipment and appliances. Coaxial cables are used for many applications that require this noise to be eliminated. For example, they can be found in the home in cable TV connections or other audiovisual connections. Coaxial cables consist of an inner conductor of radius r_i surrounded by a second, outer concentric conductor with radius r_0 (Figure 4.4.6). The space between the two is normally filled with an insulator such as polyethylene plastic. A small amount of radial leakage current occurs between the two conductors. Determine the resistance of a coaxial cable of length **L**.

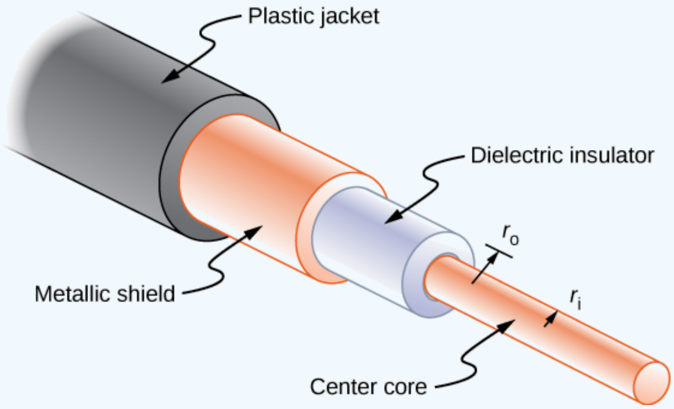


Figure 4.4.6: Coaxial cables consist of two concentric conductors separated by insulation. They are often used in cable TV or other audiovisual connections.

Strategy

We cannot use the equation $R=\rho\frac{L}{A}$ directly. Instead, we look at concentric cylindrical shells, with thickness **dr**, and integrate.

Solution

We first find an expression for dR and then integrate from r_i to r_0,

$$\begin{aligned} dR &= \frac{\rho}{A}dr \\ &= \frac{\rho}{2\pi rL}dr, \end{aligned}$$

Integrating both sides

$$\begin{aligned} R &= \int_{r_i}^{r_0} dR \\ &= \int_{r_i}^{r_0} \frac{\rho}{2\pi rL}dr \\ &= \frac{\rho}{2\pi L}\int_{r_i}^{r_0} \frac{1}{r}dr \\ &= \frac{\rho}{2\pi L}\ln\frac{r_0}{r_i}. \end{aligned}$$

Significance

The resistance of a coaxial cable depends on its length, the inner and outer radii, and the resistivity of the material separating the two conductors. Since this resistance is not infinite, a small leakage current occurs between the two conductors. This leakage current leads to the attenuation (or weakening) of the signal being sent through the cable.

Exercise 4.4.3

The resistance between the two conductors of a coaxial cable depends on the resistivity of the material separating the two conductors, the length of the cable and the inner and outer radius of the two conductor. If you are designing a coaxial cable, how does the resistance between the two conductors depend on these variables?

Answer

The longer the length, the smaller the resistance. The greater the resistivity, the higher the resistance. The larger the difference between the outer radius and the inner radius, that is, the greater the ratio between the two, the greater the resistance. If you are attempting to maximize the resistance, the choice of the values for these variables will depend on the application. For example, if the cable must be flexible, the choice of materials may be limited.

Phet: Battery-Resistor Circuit

View this simulation to see how the voltage applied and the resistance of the material the current flows through affects the current through the material. You can visualize the collisions of the electrons and the atoms of the material effect the temperature of the material.

4.5: Ohm's Law

> **Learning Objectives**
>
> By the end of this section, you will be able to:
>
> - Describe Ohm's law
> - Recognize when Ohm's law applies and when it does not

We have been discussing three electrical properties so far in this chapter: current, voltage, and resistance. It turns out that many materials exhibit a simple relationship among the values for these properties, known as Ohm's law. Many other materials do not show this relationship, so despite being called Ohm's law, it is not considered a law of nature, like Newton's laws or the laws of thermodynamics. But it is very useful for calculations involving materials that do obey Ohm's law.

Description of Ohm's Law

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is **directly proportional to the voltage applied**:

$$I \propto V.$$

This important relationship is the basis for **Ohm's law**. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law, which is to say that it is an experimentally observed phenomenon, like friction. Such a linear relationship doesn't always occur. Any material, component, or device that obeys Ohm's law, where the current through the device is proportional to the voltage applied, is known as an ohmic material or **ohmic** component. Any material or component that does not obey Ohm's law is known as a **nonohmic** material or nonohmic component.

Ohm's Experiment

In a paper published in 1827, Georg Ohm described an experiment in which he measured voltage across and current through various simple electrical circuits containing various lengths of wire. A similar experiment is shown in Figure $4.5.1$. This experiment is used to observe the current through a resistor that results from an applied voltage. In this simple circuit, a resistor is connected in series with a battery. The voltage is measured with a voltmeter, which must be placed across the resistor (in parallel with the resistor). The current is measured with an ammeter, which must be in line with the resistor (in series with the resistor).

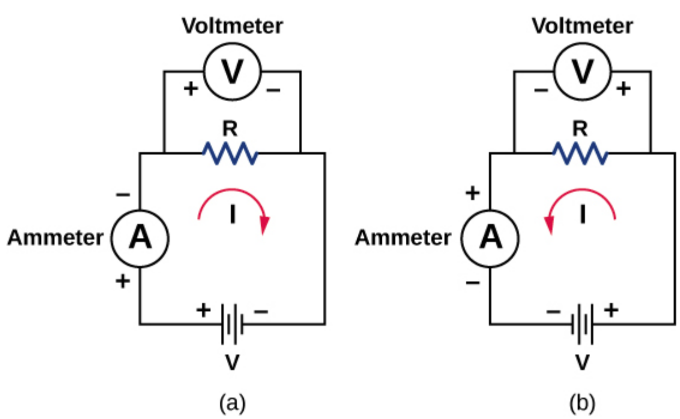


Figure $4.5.1$: The experimental set-up used to determine if a resistor is an ohmic or nonohmic device. (a) When the battery is attached, the current flows in the clockwise direction and the voltmeter and ammeter have positive readings. (b) When the leads of the battery are switched, the current flows in the counterclockwise direction and the voltmeter and ammeter have negative readings.

In this updated version of Ohm's original experiment, several measurements of the current were made for several different voltages. When the battery was hooked up as in Figure $4.5.1a$, the current flowed in the clockwise direction and the readings of the voltmeter and ammeter were positive. Does the behavior of the current change if the current flowed in the opposite direction? To get the current to flow in the opposite direction, the leads of the battery can be switched. When the leads of the battery were switched, the readings of the voltmeter and ammeter readings were negative because the current flowed in the opposite direction, in this case, counterclockwise. Results of a similar experiment are shown in Figure $4.5.2$.

I(A)	V(V)
–2.74	–10.00
–2.25	–9.00
–2.00	–8.00
–1.75	–7.00
–1.50	–6.00
–1.49	–5.00
–1.00	–4.00
–0.51	–3.00
–0.74	–2.00
–0.49	–1.00
+0.00	+0.00
+0.49	+1.00
+0.50	+2.00
+0.99	+3.00
+0.76	+4.00
+1.01	+5.00
+1.74	+6.00
+1.75	+7.00
+2.00	+8.00
+2.49	+9.00
+2.50	+10.00

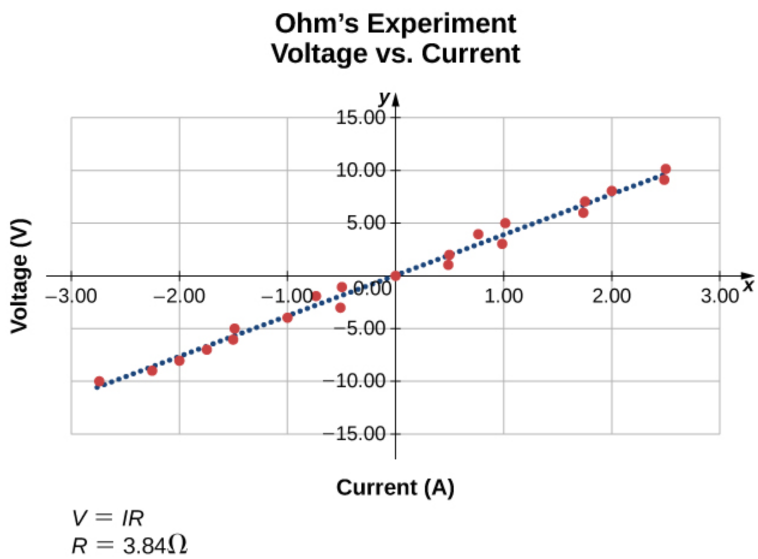


Figure $4.5.1$: A resistor is placed in a circuit with a battery. The voltage applied varies from −10.00 V to +10.00 V, increased by 1.00-V increments. A plot shows values of the voltage versus the current typical of what a casual experimenter might find.

In this experiment, the voltage applied across the resistor varies from −10.00 to +10.00 V, by increments of 1.00 V. The current through the resistor and the voltage across the resistor are measured. A plot is made of the voltage versus the current, and the result is approximately linear. The slope of the line is the resistance, or the voltage divided by the current. This result is known as **Ohm’s law**:

$$V = IR \tag{4.5.1}$$

where V is the voltage measured in volts across the object in question, I is the current measured through the object in amps, and R is the resistance in units of ohms. As stated previously, any device that shows a linear relationship between the voltage and the current is known as an ohmic device. A resistor is therefore an ohmic device.

Example $4.5.1$: Measuring Resistance

A carbon resistor at room temperature (20^oC) is attached to a 9.00-V battery and the current measured through the resistor is 3.00 mA.

a. What is the resistance of the resistor measured in ohms?
b. If the temperature of the resistor is increased to 60^oC by heating the resistor, what is the current through the resistor?

Strategy

a. The resistance can be found using Ohm’s law. Ohm’s law states that $V = IR$, so the resistance can be found using $R = V/I$.
b. First, the resistance is temperature dependent so the new resistance after the resistor has been heated can be found using $R = R_0(1 + \alpha\Delta T)$. The current can be found using Ohm’s law in the form $I = V/R$.

Solution

a. Using Ohm’s law and solving for the resistance yields the resistance at room temperature:

$$R = \frac{V}{I} = \frac{9.00\,V}{3.00 \times 10^{-3}\,A} = 3.00 \times 10^3\,\Omega = 3.00 k\Omega$$

b. The resistance at 60^oC can be found using $R = R_0(1 + \alpha\Delta T)$ where the temperature coefficient for carbon is $\alpha = -0.0005$.

$$R = R_0(1 + \alpha\Delta T) = 3.00 \times 10^3(1 - 0.0005(60^oC - 20^oC)) = 2.94\,k\Omega.$$

The current through the heated resistor is

$$I = \frac{V}{R} = \frac{9.00\,V}{2.94 \times 10^3\,\Omega} = 3.06 \times 10^{-3}\,A = 3.06\,mA.$$

Significance

A change in temperature of $40^o C$ resulted in a 2.00% change in current. This may not seem like a very great change, but changing electrical characteristics can have a strong effect on the circuits. For this reason, many electronic appliances, such as computers, contain fans to remove the heat dissipated by components in the electric circuits.

? Exercise 4.5.1

The voltage supplied to your house varies as $V(t) = V_{max}\,sin\,(2\pi\,ft)$. If a resistor is connected across this voltage, will Ohm's law $V = IR$ still be valid?

Answer

Yes, Ohm's law is still valid. At every point in time the current is equal to $I(t) = V(t)/R$, so the current is also a function of time, $I(t) = \frac{V_{max}}{R}\,sin\,(2\pi\,ft)$.

Simulation: PhET

See how Ohm's law (Equation 4.5.1) relates to a simple circuit. Adjust the voltage and resistance, and see the current change according to Ohm's law. The sizes of the symbols in the equation change to match the circuit diagram.

Nonohmic devices do not exhibit a linear relationship between the voltage and the current. One such device is the semiconducting circuit element known as a **diode**. A diode is a circuit device that allows current flow in only one direction. A diagram of a simple circuit consisting of a battery, a diode, and a resistor is shown in Figure 4.5.3. Although we do not cover the theory of the diode in this section, the diode can be tested to see if it is an ohmic or a nonohmic device.

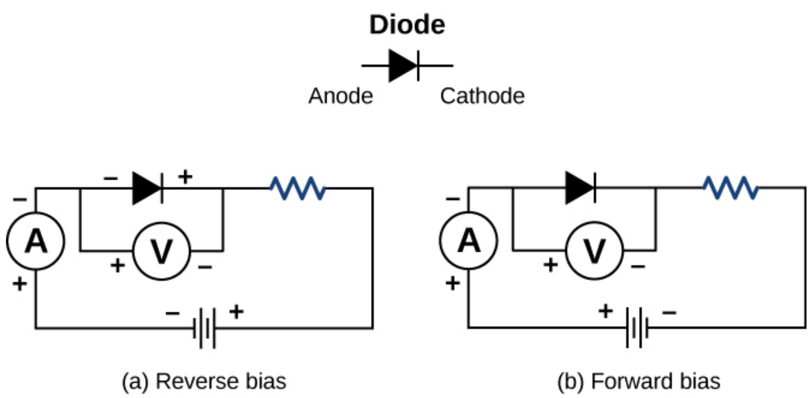


Figure 4.5.3: A diode is a semiconducting device that allows current flow only if the diode is forward biased, which means that the anode is positive and the cathode is negative.

A plot of current versus voltage is shown in Figure 4.5.4. Note that the behavior of the diode is shown as current versus voltage, whereas the resistor operation was shown as voltage versus current. A diode consists of an anode and a cathode. When the anode is at a negative potential and the cathode is at a positive potential, as shown in part (a), the diode is said to have reverse bias. With reverse bias, the diode has an extremely large resistance and there is very little current flow—essentially zero current—through the diode and the resistor. As the voltage applied to the circuit increases, the current remains essentially zero, until the voltage reaches the breakdown voltage and the diode conducts current. When the battery and the potential across the diode are reversed, making the anode positive and the cathode negative, the diode conducts and current flows through the diode if the voltage is greater than 0.7 V. The resistance of the diode is close to zero. (This is the reason for the resistor in the circuit; if it were not there, the current would become very large.) You can see from the graph in Figure 4.5.4 that the voltage and the current do not have a linear relationship. Thus, the diode is an example of a nonohmic device.

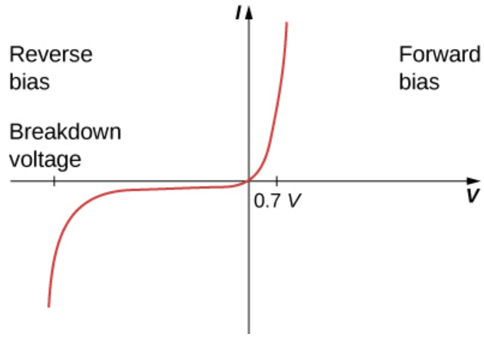


Figure 4.5.4: When the voltage across the diode is negative and small, there is very little current flow through the diode. As the voltage reaches the breakdown voltage, the diode conducts. When the voltage across the diode is positive and greater than 0.7 V (the actual voltage value depends on the diode), the diode conducts. As the voltage applied increases, the current through the diode increases, but the voltage across the diode remains approximately 0.7 V.

Ohm's law is commonly stated as $V = IR$, but originally it was stated as a microscopic view, in terms of the current density, the conductivity, and the electrical field. This microscopic view suggests the proportionality $V \propto I$ comes from the drift velocity of the free electrons in the metal that results from an applied electrical field. As stated earlier, the current density is proportional to the applied electrical field. The reformulation of Ohm's law is credited to Gustav Kirchhoff, whose name we will see again in the next chapter.

4.6: Electrical Energy and Power

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Express electrical power in terms of the voltage and the current
- Describe the power dissipated by a resistor in an electric circuit
- Calculate the energy efficiency and cost effectiveness of appliances and equipment

In an electric circuit, electrical energy is continuously converted into other forms of energy. For example, when a current flows in a conductor, electrical energy is converted into thermal energy within the conductor. The electrical field, supplied by the voltage source, accelerates the free electrons, increasing their kinetic energy for a short time. This increased kinetic energy is converted into thermal energy through collisions with the ions of the lattice structure of the conductor. Previously, we defined power as the rate at which work is done by a force measured in watts. Power can also be defined as the rate at which energy is transferred. In this section, we discuss the time rate of energy transfer, or power, in an electric circuit.

Power in Electric Circuits

Power is associated by many people with electricity. Power transmission lines might come to mind. We also think of light bulbs in terms of their power ratings in watts. What is the expression for **electric power**?

Let us compare a 25-W bulb with a 60-W bulb (Figure $4.6.1a$). The 60-W bulb glows brighter than the 25-W bulb. Although it is not shown, a 60-W light bulb is also warmer than the 25-W bulb. The heat and light is produced by from the conversion of electrical energy. The kinetic energy lost by the electrons in collisions is converted into the internal energy of the conductor and radiation. How are voltage, current, and resistance related to electric power?

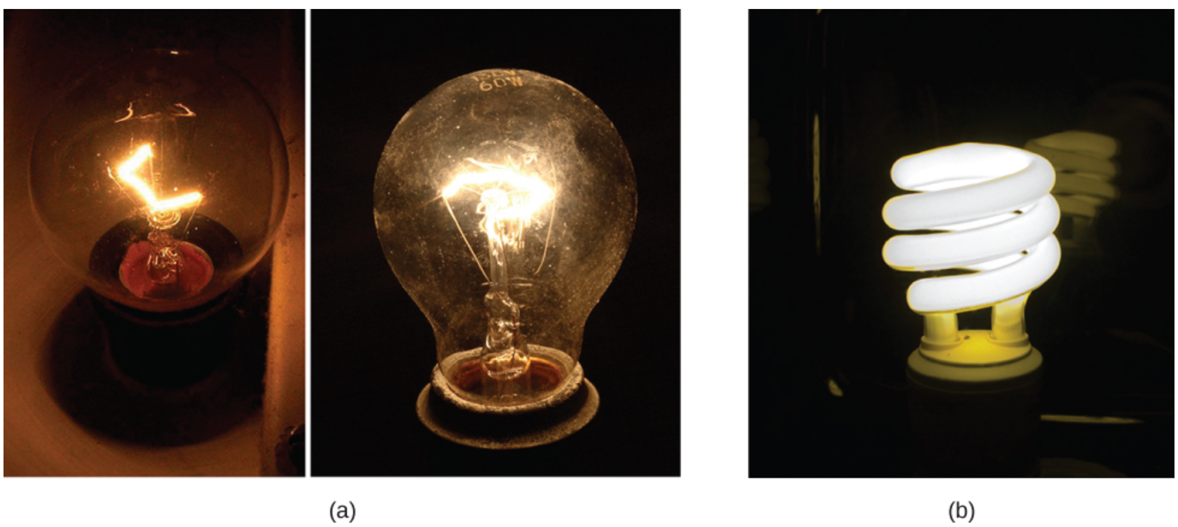


Figure $4.6.1$: (a) Pictured above are two incandescent bulbs: a 25-W bulb (left) and a 60-W bulb (right). The 60-W bulb provides a higher intensity light than the 25-W bulb. The electrical energy supplied to the light bulbs is converted into heat and light. (b) This compact fluorescent light (CFL) bulb puts out the same intensity of light as the 60-W bulb, but at 1/4 to 1/10 the input power. (credit a: modification of works by "Dickbauch"/Wikimedia Commons and Greg Westfall; credit b: modification of work by "dbgg1979"/Flickr)

To calculate electric power, consider a voltage difference existing across a material (Figure $4.6.2$). The electric potential V_1 is higher than the electric potential at V_2, and the voltage difference is negative $V = V_2 - V_1$. As discussed in Electric Potential, an electrical field exists between the two potentials, which points from the higher potential to the lower potential. Recall that the electrical potential is defined as the potential energy per charge, $V = \Delta U/q$, and the charge ΔQ loses potential energy moving through the potential difference.

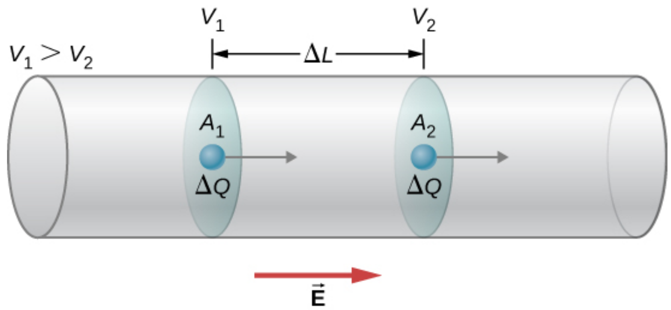


Figure 4.6.2: When there is a potential difference across a conductor, an electrical field is present that points in the direction from the higher potential to the lower potential.

If the charge is positive, the charge experiences a force due to the electrical field $\vec{F} = m\vec{a} = \Delta Q\vec{E}$. This force is necessary to keep the charge moving. This force does not act to accelerate the charge through the entire distance ΔL because of the interactions of the charge with atoms and free electrons in the material. The speed, and therefore the kinetic energy, of the charge do not increase during the entire trip across ΔL, and charge passing through area A_2 has the same drift velocity v_d as the charge that passes through area A_1. However, work is done on the charge, by the electrical field, which changes the potential energy. Since the change in the electrical potential difference is negative, the electrical field is found to be

$$E = -\frac{(V_2 - V_1)}{\Delta L} = \frac{V}{\Delta L}.$$

The work done on the charge is equal to the electric force times the length at which the force is applied,

$$W = F\Delta L = (\Delta QE)\Delta L = \left(\Delta Q\frac{V}{\Delta L}\right)\Delta L = \Delta QV = \Delta U.$$

The charge moves at a drift velocity v_d so the work done on the charge results in a loss of potential energy, but the average kinetic energy remains constant. The lost electrical potential energy appears as thermal energy in the material. On a microscopic scale, the energy transfer is due to collisions between the charge and the molecules of the material, which leads to an increase in temperature in the material. The loss of potential energy results in an increase in the temperature of the material, which is dissipated as radiation. In a resistor, it is dissipated as heat, and in a light bulb, it is dissipated as heat and light.

The power dissipated by the material as heat and light is equal to the time rate of change of the work:

$$P = IV = I(IR) = I^2R$$

or

$$P = IV = \left(\frac{V}{R}\right)V = \frac{V^2}{R}.$$

If a resistor is connected to a battery, the power dissipated as radiant energy by the wires and the resistor is equal to

$$P = IV = I^2R = \frac{V^2}{R}.$$

The power supplied from the battery is equal to current times the voltage, $P = IV$.

Definition: Electric Power

The electric power gained or lost by any device has the form

$$P = IV.$$

The power dissipated by a resistor has the form

$$P = I^2R = \frac{V^2}{R}.$$

Different insights can be gained from the three different expressions for electric power. For example, $P = V^2/R$ implies that the lower the resistance connected to a given voltage source, the greater the power delivered. Furthermore, since voltage is squared in $P = V^2/R$, the effect of applying a higher voltage is perhaps greater than expected. Thus, when the voltage is doubled to a 25-W bulb, its power nearly quadruples to about 100 W, burning it out. If the bulb's resistance remained constant, its power would be exactly 100 W, but at the higher temperature, its resistance is higher, too.

Example 4.6.1: Calculating Power in Electric Devices

A DC winch motor is rated at 20.00 A with a voltage of 115 V. When the motor is running at its maximum power, it can lift an object with a weight of 4900.00 N a distance of 10.00 m, in 30.00 s, at a constant speed.

a. What is the power consumed by the motor?
b. What is the power used in lifting the object? Ignore air resistance. (c) Assuming that the difference in the power consumed by the motor and the power used lifting the object are dissipated as heat by the resistance of the motor, estimate the resistance of the motor?

Strategy

a. The power consumed by the motor can be found using $P = IV$.
b. The power used in lifting the object at a constant speed can be found using $P = Fv$, where the speed is the distance divided by the time. The upward force supplied by the motor is equal to the weight of the object because the acceleration is zero. (c) The resistance of the motor can be found using $P = I^2R$.

Solution

1. The power consumed by the motor is equal to $P = IV$ and the current is given as 20.00 A and the voltage is 115.00 V:

$$P = IV = (20.00\,A)115.00\,V = 2300.00\,W.$$

2. The power used lifting the object is equal to $P = Fv$ where the force is equal to the weight of the object (1960 N) and the magnitude of the velocity is

$$v = \frac{10.00\,m}{30.00\,s} = 0.33\frac{m}{s}$$

$$P = Fv = (4900\,N)0.33\,m/s = 1633.33\,W.$$

3. The difference in the power equals $2300.00\,W - 1633.33\,W = 666.67\,W$and the resistance can be found using $P = I^2R$:

$$R = \frac{P}{I^2} = \frac{666.67\,W}{(20.00\,A)^2} = 1.67\,\Omega.$$

Significance The resistance of the motor is quite small. The resistance of the motor is due to many windings of copper wire. The power dissipated by the motor can be significant since the thermal power dissipated by the motor is proportional to the square of the current $(P = I^2R)$.

Exercise 4.6.1

Electric motors have a reasonably high efficiency. A 100-hp motor can have an efficiency of 90% and a 1-hp motor can have an efficiency of 80%. Why is it important to use high-performance motors?

Answer

Even though electric motors are highly efficient 10–20% of the power consumed is wasted, not being used for doing useful work. Most of the 10–20% of the power lost is transferred into heat dissipated by the copper wires used to make the coils of the motor. This heat adds to the heat of the environment and adds to the demand on power plants providing the power. The demand on the power plant can lead to increased greenhouse gases, particularly if the power plant uses coal or gas as fuel.

A fuse (Figure 4.6.3) is a device that protects a circuit from currents that are too high. A fuse is basically a short piece of wire between two contacts. As we have seen, when a current is running through a conductor, the kinetic energy of the charge carriers is converted into thermal energy in the conductor. The piece of wire in the fuse is under tension and has a low melting point. The wire is designed to heat up and break at the rated current. The fuse is destroyed and must be replaced, but it protects the rest of the circuit. Fuses act quickly, but there is a small time delay while the wire heats up and breaks.

Figure $4.6.3$: A fuse consists of a piece of wire between two contacts. When a current passes through the wire that is greater than the rated current, the wire melts, breaking the connection. Pictured is a "blown" fuse where the wire broke protecting a circuit (credit: modification of work by "Shardayyy"/Flickr).

Circuit breakers are also rated for a maximum current, and open to protect the circuit, but can be reset. Circuit breakers react much faster. The operation of circuit breakers is not within the scope of this chapter and will be discussed in later chapters. Another method of protecting equipment and people is the ground fault circuit interrupter (GFCI), which is common in bathrooms and kitchens. The GFCI outlets respond very quickly to changes in current. These outlets open when there is a change in magnetic field produced by current-carrying conductors, which is also beyond the scope of this chapter and is covered in a later chapter.

The Cost of Electricity

The more electric appliances you use and the longer they are left on, the higher your electric bill. This familiar fact is based on the relationship between energy and power. You pay for the energy used. Since $P = \frac{dE}{dt}$, we see that

$$E = \int P dt$$

is the energy used by a device using power **P** for a time interval **t**. If power is delivered at a constant rate, then then the energy can be found by $E = Pt$. For example, the more light bulbs burning, the greater **P** used; the longer they are on, the greater **t** is.

The energy unit on electric bills is the kilowatt-hour $(kW \cdot h)$, consistent with the relationship $E = Pt$. It is easy to estimate the cost of operating electrical appliances if you have some idea of their power consumption rate in watts or kilowatts, the time they are on in hours, and the cost per kilowatt-hour for your electric utility. Kilowatt-hours, like all other specialized energy units such as food calories, can be converted into joules. You can prove to yourself that $1\ kW \cdot h = 3.6 \times 10^6\ J$.

The electrical energy (**E**) used can be reduced either by reducing the time of use or by reducing the power consumption of that appliance or fixture. This not only reduces the cost but also results in a reduced impact on the environment. Improvements to lighting are some of the fastest ways to reduce the electrical energy used in a home or business. About 20% of a home's use of energy goes to lighting, and the number for commercial establishments is closer to 40%. Fluorescent lights are about four times more efficient than incandescent lights—this is true for both the long tubes and the compact fluorescent lights (CFLs), e.g., Figure $4.6.1b$. Thus, a 60-W incandescent bulb can be replaced by a 15-W CFL, which has the same brightness and color. CFLs have a bent tube inside a globe or a spiral-shaped tube, all connected to a standard screw-in base that fits standard incandescent light sockets. (Original problems with color, flicker, shape, and high initial investment for CFLs have been addressed in recent years.)

The heat transfer from these CFLs is less, and they last up to 10 times longer than incandescent bulbs. The significance of an investment in such bulbs is addressed in the next example. New white LED lights (which are clusters of small LED bulbs) are even more efficient (twice that of CFLs) and last five times longer than CFLs.

Example \(\PageIndex{2}\): Calculating the Cost Effectiveness of LED Bulb

The typical replacement for a 100-W incandescent bulb is a 20-W LED bulb. The 20-W LED bulb can provide the same amount of light output as the 100-W incandescent light bulb. What is the cost savings for using the LED bulb in place of the incandescent bulb for one year, assuming $0.10 per kilowatt-hour is the average energy rate charged by the power company? Assume that the bulb is turned on for three hours a day.

Strategy

a. Calculate the energy used during the year for each bulb, using $E = Pt$.
b. Multiply the energy by the cost.

Solution

a. Calculate the power for each bulb.

$$E_{Incandescent} = Pt = 100\,W\left(\frac{1\,kW}{1000\,W}\right)\left(\frac{3\,h}{day}\right)(365\,days) = 109.5\,kW\cdot h$$

$$E_{LED} = Pt = 20\,W\left(\frac{1\,kW}{1000\,W}\right)\left(\frac{3\,h}{day}\right)(365\,days) = 21.9\,kW\cdot h$$

b. Calculate the cost for each.

$$cost_{Incandescent} = 109.5\,kW\cdot h\left(\frac{\$0.10}{kW\cdot h}\right) = \$10.95$$

$$cost_{LED} = 21.90\,kW\cdot h\left(\frac{\$0.10}{kW\cdot h}\right) = \$2.19$$

Significance

A LED bulb uses 80% less energy than the incandescent bulb, saving $8.76 over the incandescent bulb for one year. The LED bulb can cost $20.00 and the 100-W incandescent bulb can cost $0.75, which should be calculated into the computation. A typical lifespan of an incandescent bulb is 1200 hours and is 50,000 hours for the LED bulb. The incandescent bulb would last 1.08 years at 3 hours a day and the LED bulb would last 45.66 years. The initial cost of the LED bulb is high, but the cost to the home owner will be $0.69 for the incandescent bulbs versus $0.44 for the LED bulbs per year. (Note that the LED bulbs are coming down in price.) The cost savings per year is approximately $8.50, and that is just for one bulb.

? Exercise 4.6.2

Is the efficiency of the various light bulbs the only consideration when comparing the various light bulbs?

Answer

No, the efficiency is a very important consideration of the light bulbs, but there are many other considerations. As mentioned above, the cost of the bulbs and the life span of the bulbs are important considerations. For example, CFL bulbs contain mercury, a neurotoxin, and must be disposed of as hazardous waste. When replacing incandescent bulbs that are being controlled by a dimmer switch with LED, the dimmer switch may need to be replaced. The dimmer switches for LED lights are comparably priced to the incandescent light switches, but this is an initial cost which should be considered. The spectrum of light should also be considered, but there is a broad range of color temperatures available, so you should be able to find one that fits your needs. None of these considerations mentioned are meant to discourage the use of LED or CFL light bulbs, but they are considerations.

Changing light bulbs from incandescent bulbs to CFL or LED bulbs is a simple way to reduce energy consumption in homes and commercial sites. CFL bulbs operate with a much different mechanism than do incandescent lights. The mechanism is complex and beyond the scope of this chapter, but here is a very general description of the mechanism. CFL bulbs contain argon and mercury vapor housed within a spiral-shaped tube. The CFL bulbs use a “ballast” that increases the voltage used by the CFL bulb. The ballast produce an electrical current, which passes through the gas mixture and excites the gas molecules. The excited gas molecules produce ultraviolet (UV) light, which in turn stimulates the fluorescent coating on the inside of the tube. This coating fluoresces in the visible spectrum, emitting visible light. Traditional fluorescent tubes and CFL bulbs had a short time delay of up to a few seconds while the mixture was being “warmed up” and the molecules reached an excited state. It should be noted that these bulbs do contain mercury, which is poisonous, but if the bulb is broken, the mercury is never released. Even if the bulb is broken, the mercury tends to remain in the fluorescent coating. The amount is also quite small and the advantage of the energy saving may outweigh the disadvantage of using mercury.

The CFL light bulbs are being replaced with LED light bulbs, where LED stands for “light-emitting diode.” The diode was briefly discussed as a nonohmic device, made of semiconducting material, which essentially permits current flow in one direction. LEDs are a special type of diode made of semiconducting materials infused with impurities in combinations and concentrations that enable the extra energy from the movement of the electrons during electrical excitation to be converted into visible light. Semiconducting devices will be explained in greater detail in Condensed Matter Physics.

Commercial LEDs are quickly becoming the standard for commercial and residential lighting, replacing incandescent and CFL bulbs. They are designed for the visible spectrum and are constructed from gallium doped with arsenic and phosphorous atoms. The color emitted from an LED depends on the materials used in the semiconductor and the current. In the early years of LED development, small LEDs found on circuit boards were red, green, and yellow, but LED light bulbs can now be programmed to produce millions of colors of light as well as many different hues of white light.

Comparison of Incandescent, CFL, and LED Light Bulbs

The energy savings can be significant when replacing an incandescent light bulb or a CFL light bulb with an LED light. Light bulbs are rated by the amount of power that the bulb consumes, and the amount of light output is measured in lumens. The lumen (lm) is the SI -derived unit of luminous flux and is a measure of the total quantity of visible light emitted by a source. A 60-W incandescent light bulb can be replaced with a 13- to 15-W CFL bulb or a 6- to 8-W LED bulb, all three of which have a light output of approximately 800 lm. A table of light output for some commonly used light bulbs appears in Table $4.6.1$.

The life spans of the three types of bulbs are significantly different. An LED bulb has a life span of 50,000 hours, whereas the CFL has a lifespan of 8000 hours and the incandescent lasts a mere 1200 hours. The LED bulb is the most durable, easily withstanding rough treatment such as jarring and bumping. The incandescent light bulb has little tolerance to the same treatment since the filament and glass can easily break. The CFL bulb is also less durable than the LED bulb because of its glass construction. The amount of heat emitted is 3.4 btu/h for the 8-W LED bulb, 85 btu/h for the 60-W incandescent bulb, and 30 btu/h for the CFL bulb. As mentioned earlier, a major drawback of the CFL bulb is that it contains mercury, a neurotoxin, and must be disposed of as hazardous waste. From these data, it is easy to understand why the LED light bulb is quickly becoming the standard in lighting.

Table $4.6.1$: Light Output of LED, Incandescent, and CFL Light Bulbs

Light Output (lumens)	LED Light Bulbs (watts)	Incandescent Light Bulbs (watts)	CFL Light Bulbs (watts)
450	4−5	40	9−13
800	6−8	60	13−15
1100	9−13	75	18−25
1600	16−20	100	23−30
2600	25−28	150	30−55

Summary of Relationships

In this chapter, we have discussed relationships between voltages, current, resistance, and power. Figure $4.6.4$ shows a summary of the relationships between these measurable quantities for ohmic devices. (Recall that ohmic devices follow Ohm’s law $V = IR$.) For example, if you need to calculate the power, use the pink section, which shows that $P = VI$, $P = \frac{V^2}{R}$, and $P = I^2R$.

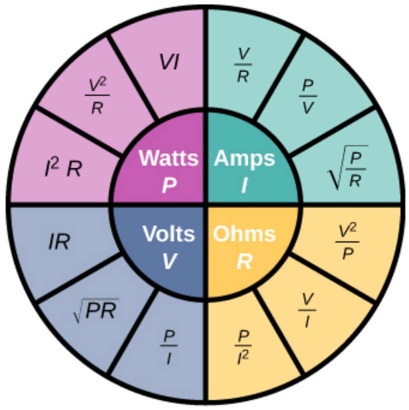


Figure 4.6.4: This circle shows a summary of the equations for the relationships between power, current, voltage, and resistance.

Which equation you use depends on what values you are given, or you measure. For example if you are given the current and the resistance, use $P = I^2R$. Although all the possible combinations may seem overwhelming, don't forget that they all are combinations of just two equations, Ohm's law $(V = IR)$ and power $(P = IV)$.

4.7: Superconductors

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the phenomenon of superconductivity
- List applications of superconductivity

Touch the power supply of your laptop computer or some other device. It probably feels slightly warm. That heat is an unwanted byproduct of the process of converting household electric power into a current that can be used by your device. Although electric power is reasonably efficient, other losses are associated with it. As discussed in the section on power and energy, transmission of electric power produces I^2R line losses. These line losses exist whether the power is generated from conventional power plants (using coal, oil, or gas), nuclear plants, solar plants, hydroelectric plants, or wind farms. These losses can be reduced, but not eliminated, by transmitting using a higher voltage. It would be wonderful if these line losses could be eliminated, but that would require transmission lines that have zero resistance. In a world that has a global interest in not wasting energy, the reduction or elimination of this unwanted thermal energy would be a significant achievement. Is this possible?

The Resistance of Mercury

In 1911, Heike **Kamerlingh Onnes** of Leiden University, a Dutch physicist, was looking at the temperature dependence of the resistance of the element mercury. He cooled the sample of mercury and noticed the familiar behavior of a linear dependence of resistance on temperature; as the temperature decreased, the resistance decreased. Kamerlingh Onnes continued to cool the sample of mercury, using liquid helium. As the temperature approached $4.2\,K\,(-269.2^oC)$, the resistance abruptly went to zero (Figure $4.7.1$). This temperature is known as the **critical temperature** T_c for mercury. The sample of mercury entered into a phase where the resistance was absolutely zero. This phenomenon is known as **superconductivity**. (**Note:** If you connect the leads of a three-digit ohmmeter across a conductor, the reading commonly shows up as $0.00\,\Omega$. The resistance of the conductor is not actually zero, it is less than $0.01\,\Omega$.) There are various methods to measure very small resistances, such as the four-point method, but an ohmmeter is not an acceptable method to use for testing resistance in superconductivity.

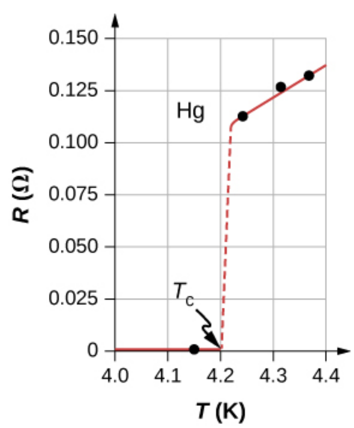


Figure $4.7.1$: The resistance of a sample of mercury is zero at very low temperatures—it is a superconductor up to the temperature of about 4.2 K. Above that critical temperature, its resistance makes a sudden jump and then increases nearly linearly with temperature.

Other Superconducting Materials

As research continued, several other materials were found to enter a superconducting phase, when the temperature reached near absolute zero. In 1941, an alloy of niobium-nitride was found that could become superconducting at $T_c = 16\,K\,(-257^oC)$ and in 1953, vanadium-silicon was found to become superconductive at $T_c = 17.5\,K\,(-255.7^oC)$. The temperatures for the transition into superconductivity were slowly creeping higher. Strangely, many materials that make good conductors, such as copper, silver, and gold, do not exhibit superconductivity. Imagine the energy savings if transmission lines for electric power-generating stations could be made to be superconducting at temperatures near room temperature! A resistance of zero ohms means no I^2R losses and a great boost to reducing energy consumption. The problem is that $T_c = 17.5\,K$ is still very cold and in the range of liquid helium temperatures. At this temperature, it is not cost effective to transmit electrical energy because of the cooling requirements.

A large jump was seen in 1986, when a team of researchers, headed by Dr. Ching Wu Chu of Houston University, fabricated a brittle, ceramic compound with a transition temperature of $T_c = 92\,K\,(-181^oC)$. The ceramic material, composed of yttrium barium copper oxide (YBCO), was an insulator at room temperature. Although this temperature still seems quite cold, it is near the boiling point of liquid nitrogen, a liquid commonly used in refrigeration. You may have noticed refrigerated trucks traveling down the highway labeled as "Liquid Nitrogen Cooled."

YBCO ceramic is a material that could be useful for transmitting electrical energy because the cost saving of reducing the I^2R losses are larger than the cost of cooling the superconducting cable, making it financially feasible. There were and are many engineering problems to overcome. For example, unlike traditional electrical cables, which are flexible and have a decent tensile strength, ceramics are brittle and would break rather than stretch under pressure. Processes that are rather simple with traditional cables, such as making connections, become difficult when working with ceramics. The problems are difficult and complex, and material scientists and engineers are coming up with innovative solutions.

An interesting consequence of the resistance going to zero is that once a current is established in a superconductor, it persists without an applied voltage source. Current loops in a superconductor have been set up and the current loops have been observed to persist for years without decaying.

Zero resistance is not the only interesting phenomenon that occurs as the materials reach their transition temperatures. A second effect is the exclusion of magnetic fields. This is known as the **Meissner effect** (Figure $4.7.2$). A light, permanent magnet placed over a superconducting sample will levitate in a stable position above the superconductor. High-speed trains have been developed that levitate on strong superconducting magnets, eliminating the friction normally experienced between the train and the tracks. In Japan, the Yamanashi Maglev test line opened on April 3, 1997. In April 2015, the MLX01 test vehicle attained a speed of 374 mph (603 km/h).

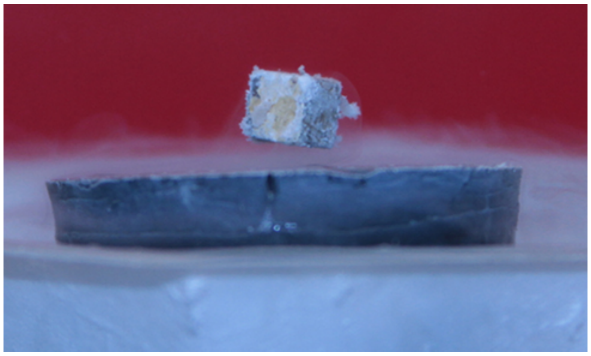


Figure $4.7.1$: A small, strong magnet levitates over a superconductor cooled to liquid nitrogen temperature. The magnet levitates because the superconductor excludes magnetic fields.

Table $4.7.1$ shows a select list of elements, compounds, and high-temperature superconductors, along with the critical temperatures for which they become superconducting. Each section is sorted from the highest critical temperature to the lowest. Also listed is the critical magnetic field for some of the materials. This is the strength of the magnetic field that destroys superconductivity. Finally, the type of the superconductor is listed.

There are two types of superconductors. There are 30 pure metals that exhibit zero resistivity below their critical temperature and exhibit the Meissner effect, the property of excluding magnetic fields from the interior of the superconductor while the superconductor is at a temperature below the critical temperature. These metals are called Type I superconductors. The superconductivity exists only below their critical temperatures and below a critical magnetic field strength. Type I superconductors are well described by the BCS theory (described next). Type I superconductors have limited practical applications because the strength of the critical magnetic field needed to destroy the superconductivity is quite low.

Type II superconductors are found to have much higher critical magnetic fields and therefore can carry much higher current densities while remaining in the superconducting state. A collection of various ceramics containing barium-copper-oxide have much higher critical temperatures for the transition into a superconducting state. Superconducting materials that belong to this subcategory of the Type II superconductors are often categorized as high-temperature superconductors.

Introduction to BCS Theory

Type I superconductors, along with some Type II superconductors can be modeled using the **BCS theory**, proposed by John **Bardeen**, Leon **Cooper**, and Robert **Schrieffer**. Although the theory is beyond the scope of this chapter, a short summary of the theory is provided here. (More detail is provided in Condensed Matter Physics.) The theory considers pairs of electrons and how they are coupled together through lattice-vibration interactions. Through the interactions with the crystalline lattice, electrons near the Fermi energy level feel a small attractive force and form pairs (**Cooper pairs**), and the coupling is known as a phonon interaction. Single electrons are fermions, which are particles that obey the Pauli exclusion principle. The Pauli exclusion principle in quantum mechanics states that two identical fermions (particles with half-integer spin) cannot occupy the same quantum state simultaneously. Each electron has four quantum numbers $(n,\ \ell,\ m_\ell,\ m_s)$. The principal quantum number (n) describes the energy of the electron, the orbital angular momentum quantum number (ℓ) indicates the most probable distance from the nucleus, the magnetic quantum number m_ℓ describes the energy levels in the subshell, and the electron spin quantum number m_s describes the orientation of the spin of the electron, either up or down. As the material enters a superconducting state, pairs of electrons act more like bosons, which can condense into the same energy level and need not obey the Pauli exclusion principle. The electron pairs have a slightly lower energy and leave an energy gap above them on the order of 0.001 eV. This energy gap inhibits collision interactions that lead to ordinary resistivity. When the material is below the critical temperature, the thermal energy is less than the band gap and the material exhibits zero resistivity.

Table 4.7.1: Superconductor Critical Temperatures

Material	Symbol or Formula	Critical Temperature $T_c(K)$	Critical Magnetic Field $\boldsymbol{H_c(T)}$	Type
Elements				
Lead	Pb	7.19	0.08	I
Lanthanum	La	(α) 4.90 - (β) 6.30		I
Tantalum	Ta	4.48	0.09	I
Mercury	Hg	(α) 4.15 - (β) 3.95	0.04	I
Tin	Sn	3.72	0.03	I
Indium	In	3.40	0.03	I
Thallium	Tl	2.39	0.03	I
Rhenium	Re	2.40	0.03	I
Thorium	Th	1.37	0.013	I
Protactinium	Pa	1.40		I
Aluminum	Al	1.20	0.01	I
Gallium	Ga	1.10	0.005	I
Zinc	Zn	0.86	0.014	I
Titanium	Ti	0.39	0.01	I
Uranium	U	(α) 0.68 - (β) 1.80		I
Cadmium	Cd	11.4	4.00	I
Compounds				
Niobium-germanium	Nb_3Ge	23.20	37.00	II
Niobium-tin	Nb_3Sn	18.30	30.00	II
Niobium-nitrite	NbN	16.00		II
Niobium-titanium	NbTi	10.00	15.00	II
High-Temperature Oxides				

Material	Symbol or Formula	Critical Temperature $T_c(K)$	Critical Magnetic Field $H_c(T)$	Type
	$HgBa_2CaCu_2O_8$	134.00		II
	$Ti_2Ba_2Ca_2Cu_3O_{10}$	125.00		II
	$YBa_2Cu_3O_7$	92.00	120.00	II

Applications of Superconductors

Superconductors can be used to make superconducting magnets. These magnets are 10 times stronger than the strongest electromagnets. These magnets are currently in use in magnetic resonance imaging (MRI), which produces high-quality images of the body interior without dangerous radiation.

Another interesting application of superconductivity is the **SQUID** (superconducting quantum interference device). A SQUID is a very sensitive magnetometer used to measure extremely subtle magnetic fields. The operation of the SQUID is based on superconducting loops containing Josephson junctions. A **Josephson junction** is the result of a theoretical prediction made by B. D. Josephson in an article published in 1962. In the article, Josephson described how a supercurrent can flow between two pieces of superconductor separated by a thin layer of insulator. This phenomenon is now called the **Josephson effect**. The SQUID consists of a superconducting current loop containing two Josephson junctions, as shown in Figure $4.7.3$. When the loop is placed in even a very weak magnetic field, there is an interference effect that depends on the strength of the magnetic field.

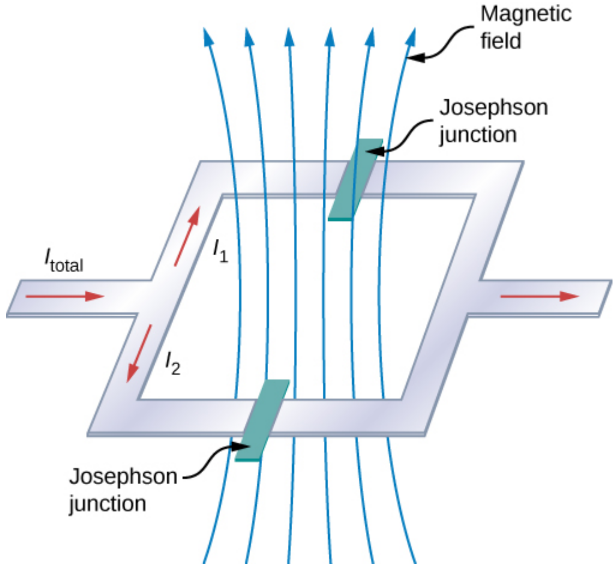


Figure $4.7.3$: The SQUID (superconducting quantum interference device) uses a superconducting current loop and two Josephson junctions to detect magnetic fields as low as 10^{-14} (Earth's magnet field is on the order of $0.3 \times 10^{-5}T$).

Superconductivity is a fascinating and useful phenomenon. At critical temperatures near the boiling point of liquid nitrogen, superconductivity has special applications in MRIs, particle accelerators, and high-speed trains. Will we reach a state where we can have materials enter the superconducting phase at near room temperatures? It seems a long way off, but if scientists in 1911 were asked if we would reach liquid-nitrogen temperatures with a ceramic, they might have thought it implausible.

CHAPTER OVERVIEW

5: Direct Current Circuits

Topic hierarchy

5.1: Prelude to Direct-Current Circuits

In the preceding few chapters, we discussed electric components, including capacitors, resistors, and diodes. In this chapter, we use these electric components in circuits. A circuit is a collection of electrical components connected to accomplish a specific task. Figure $5.1.1$ shows an amplifier circuit, which takes a small-amplitude signal and amplifies it to power the speakers in earbuds. Although the circuit looks complex, it actually consists of a set of series, parallel, and series-parallel circuits. The second section of this chapter covers the analysis of series and parallel circuits that consist of resistors. Later in this chapter, we introduce the basic equations and techniques to analyze any circuit, including those that are not reducible through simplifying parallel and series elements. But first, we need to understand how to power a circuit.

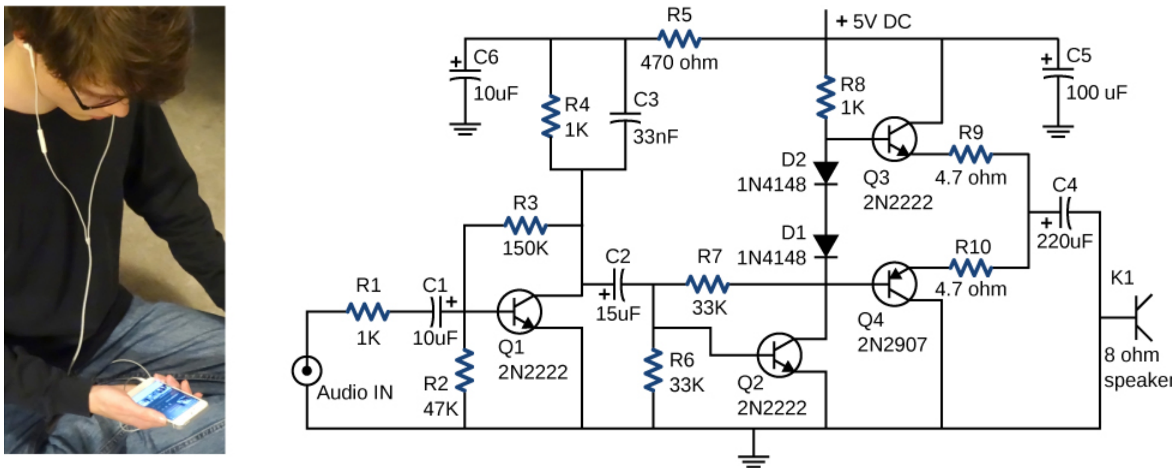


Figure $5.1.1$: This circuit shown is used to amplify small signals and power the earbud speakers attached to a cellular phone. This circuit's components include resistors, capacitors, and diodes, all of which have been covered in previous chapters, as well as transistors, which are semi-conducting devices covered in Condensed Matter Physics. Circuits using similar components are found in all types of equipment and appliances you encounter in everyday life, such as alarm clocks, televisions, computers, and refrigerators. (credit: Jane Whitney)

5.2: Electromotive Force

Learning Objectives

By the end of the section, you will be able to:

- Describe the electromotive force (emf) and the internal resistance of a battery
- Explain the basic operation of a battery

If you forget to turn off your car lights, they slowly dim as the battery runs down. Why don't they suddenly blink off when the battery's energy is gone? Their gradual dimming implies that the battery output voltage decreases as the battery is depleted. The reason for the decrease in output voltage for depleted batteries is that all voltage sources have two fundamental parts—a source of electrical energy and an internal resistance. In this section, we examine the energy source and the internal resistance.

Introduction to Electromotive Force

Voltage has many sources, a few of which are shown in Figure $5.2.2$. All such devices create a **potential difference** and can supply current if connected to a circuit. A special type of potential difference is known as **electromotive force (emf)**. The emf is not a force at all, but the term 'electromotive force' is used for historical reasons. It was coined by Alessandro Volta in the 1800s, when he invented the first battery, also known as the **voltaic pile**. Because the electromotive force is not a force, it is common to refer to these sources simply as sources of emf (pronounced as the letters "ee-em-eff"), instead of sources of electromotive force.

Figure $5.2.1$: A variety of voltage sources. (a) The Brazos Wind Farm in Fluvanna, Texas; (b) the Krasnoyarsk Dam in Russia; (c) a solar farm; (d) a group of nickel metal hydride batteries. The voltage output of each device depends on its construction and load. The voltage output equals emf only if there is no load. (credit a: modification of work by "Leaflet"/Wikimedia Commons; credit b: modification of work by Alex Polezhaev; credit c: modification of work by US Department of Energy; credit d: modification of work by Tiaa Monto)

If the electromotive force is not a force at all, then what is the emf and what is a source of emf? To answer these questions, consider a simple circuit of a 12-V lamp attached to a 12-V battery, as shown in Figure $5.2.2$. The **battery** can be modeled as a two-terminal device that keeps one terminal at a higher electric potential than the second terminal. The higher electric potential is sometimes called the positive terminal and is labeled with a plus sign. The lower-potential terminal is sometimes called the negative terminal and labeled with a minus sign. This is the source of the emf.

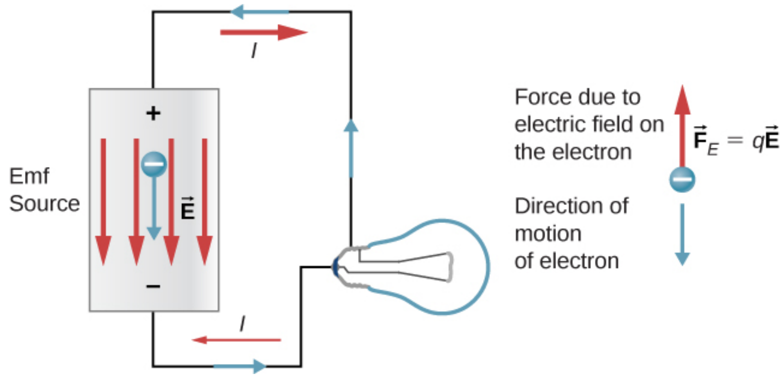


Figure $5.2.2$: A source of emf maintains one terminal at a higher electric potential than the other terminal, acting as a source of current in a circuit.

When the emf source is not connected to the lamp, there is no net flow of charge within the emf source. Once the battery is connected to the lamp, charges flow from one terminal of the battery, through the lamp (causing the lamp to light), and back to the other terminal of the battery. If we consider positive (conventional) current flow, positive charges leave the positive terminal, travel through the lamp, and enter the negative terminal.

Positive current flow is useful for most of the circuit analysis in this chapter, but in metallic wires and resistors, electrons contribute the most to current, flowing in the opposite direction of positive current flow. Therefore, it is more realistic to consider the movement of electrons for the analysis of the circuit in Figure $5.2.2$. The electrons leave the negative terminal, travel through the lamp, and return to the positive terminal. In order for the emf source to maintain the potential difference between the two terminals, negative charges (electrons) must be moved from the positive terminal to the negative terminal. The emf source acts as a charge pump, moving negative charges from the positive terminal to the negative terminal to maintain the potential difference. This increases the potential energy of the charges and, therefore, the electric potential of the charges.

The force on the negative charge from the electric field is in the opposite direction of the electric field, as shown in Figure $5.2.2$. In order for the negative charges to be moved to the negative terminal, work must be done on the negative charges. This requires energy, which comes from chemical reactions in the battery. The potential is kept high on the positive terminal and low on the negative terminal to maintain the potential difference between the two terminals. The emf is equal to the work done on the charge per unit charge $\left(\epsilon = \frac{dW}{dq}\right)$ when there is no current flowing. Since the unit for work is the joule and the unit for charge is the coulomb, the unit for emf is the volt $(1\,V = 1\,J/C)$.

The **terminal voltage** $V_{terminal}$ of a battery is voltage measured across the terminals of the battery when there is no load connected to the terminal. An ideal battery is an emf source that maintains a constant terminal voltage, independent of the current between the two terminals. An ideal battery has no internal resistance, and the terminal voltage is equal to the emf of the battery. In the next section, we will show that a real battery does have internal resistance and the terminal voltage is always less than the emf of the battery.

The Origin of Battery Potential

The combination of chemicals and the makeup of the terminals in a battery determine its emf. The **lead acid battery** used in cars and other vehicles is one of the most common combinations of chemicals. Figure $5.2.3$ shows a single cell (one of six) of this battery. The cathode (positive) terminal of the cell is connected to a lead oxide plate, whereas the anode (negative) terminal is connected to a lead plate. Both plates are immersed in sulfuric acid, the electrolyte for the system.

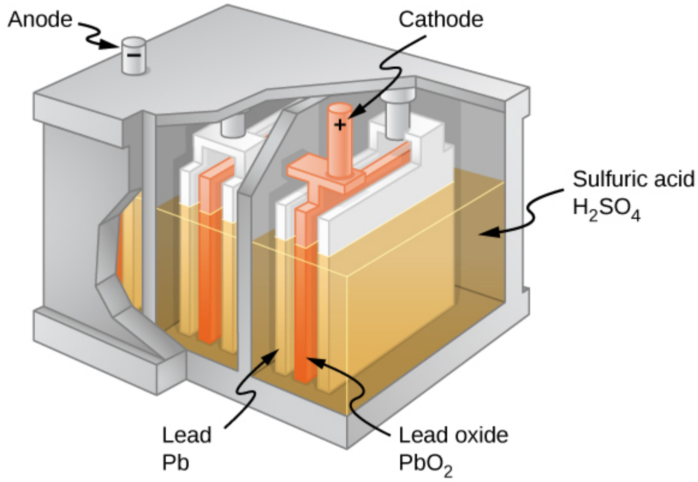


Figure 5.2.3: Chemical reactions in a lead-acid cell separate charge, sending negative charge to the anode, which is connected to the lead plates. The lead oxide plates are connected to the positive or cathode terminal of the cell. Sulfuric acid conducts the charge, as well as participates in the chemical reaction.

Knowing a little about how the chemicals in a lead-acid battery interact helps in understanding the potential created by the battery. Figure 5.2.4 shows the result of a single chemical reaction. Two electrons are placed on the **anode**, making it negative, provided that the **cathode** supplies two electrons. This leaves the cathode positively charged, because it has lost two electrons. In short, a separation of charge has been driven by a chemical reaction.

Note that the reaction does not take place unless there is a complete circuit to allow two electrons to be supplied to the cathode. Under many circumstances, these electrons come from the anode, flow through a resistance, and return to the cathode. Note also that since the chemical reactions involve substances with resistance, it is not possible to create the emf without an internal resistance.

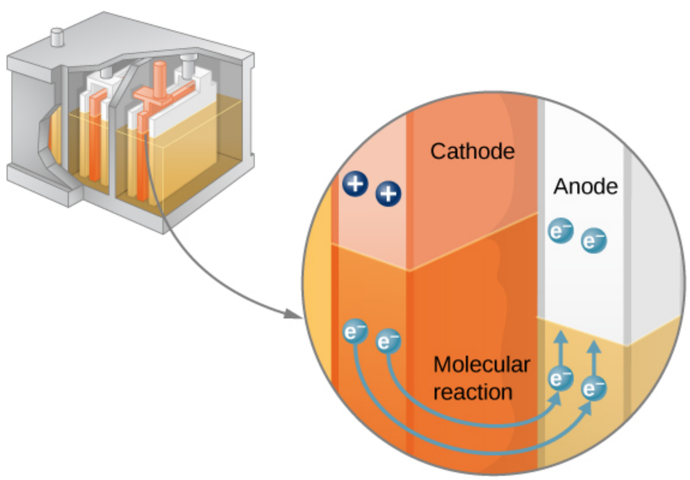


Figure 5.2.4: In a lead-acid battery, two electrons are forced onto the anode of a cell, and two electrons are removed from the cathode of the cell. The chemical reaction in a lead-acid battery places two electrons on the anode and removes two from the cathode. It requires a closed circuit to proceed, since the two electrons must be supplied to the cathode.

Internal Resistance and Terminal Voltage

The amount of resistance to the flow of current within the voltage source is called the **internal resistance**. The internal resistance **r** of a battery can behave in complex ways. It generally increases as a battery is depleted, due to the oxidation of the plates or the reduction of the acidity of the electrolyte. However, internal resistance may also depend on the magnitude and direction of the current through a voltage source, its temperature, and even its history. The internal resistance of rechargeable nickel-cadmium cells, for example, depends on how many times and how deeply they have been depleted. A simple model for a battery consists of an idealized emf source ϵ and an internal resistance **r** (Figure 5.2.5).

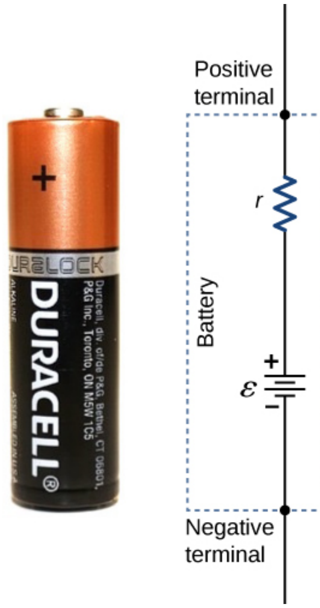


Figure $5.2.5$: A battery can be modeled as an idealized emf (ϵ) with an internal resistance (**r**). The terminal voltage of the battery is $V_{terminal} = \epsilon - Ir$.

Suppose an external resistor, known as the load resistance **R**, is connected to a voltage source such as a battery, as in Figure $5.2.6$. The figure shows a model of a battery with an emf , an internal resistance **r**, and a load resistor **R** connected across its terminals. Using conventional current flow, positive charges leave the positive terminal of the battery, travel through the resistor, and return to the negative terminal of the battery. The terminal voltage of the battery depends on the emf, the internal resistance, and the current, and is equal to

> **Note**
>
> $$V_{terminal} = \epsilon - Ir$$

For a given emf and internal resistance, the terminal voltage decreases as the current increases due to the potential drop **Ir** of the internal resistance.

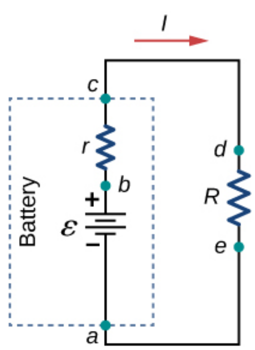


Figure $5.2.6$: Schematic of a voltage source and its load resistor **R**. Since the internal resistance **r** is in series with the load, it can significantly affect the terminal voltage and the current delivered to the load.

A graph of the potential difference across each element the circuit is shown in Figure $5.2.7$. A current **I** runs through the circuit, and the potential drop across the internal resistor is equal to **Ir**. The terminal voltage is equal to $\epsilon - Ir$, which is equal to the **potential drop** across the load resistor $IR = \epsilon - Ir$. As with potential energy, it is the change in voltage that is important. When the term “voltage” is used, we assume that it is actually the change in the potential, or ΔV. However, Δ is often omitted for convenience.

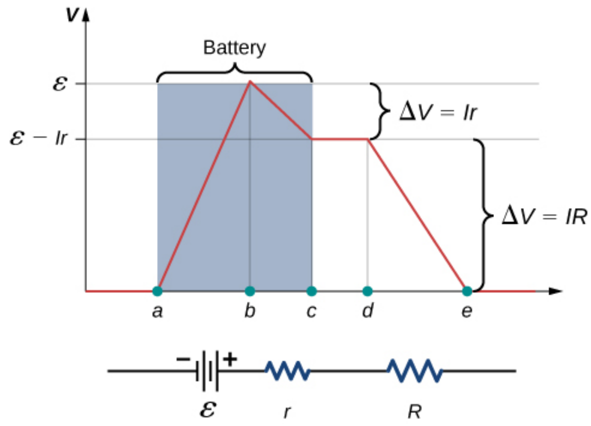


Figure 5.2.7: A graph of the voltage through the circuit of a battery and a load resistance. The electric potential increases the emf of the battery due to the chemical reactions doing work on the charges. There is a decrease in the electric potential in the battery due to the internal resistance. The potential decreases due to the internal resistance $-Ir$, making the terminal voltage of the battery equal to $(\epsilon - Ir)$. The voltage then decreases by (**IR**). The current is equal to $I = \frac{\epsilon}{r+R}$.

The current through the load resistor is $I = \frac{\epsilon}{r+R}$. We see from this expression that the smaller the internal resistance **r**, the greater the current the voltage source supplies to its load **R**. As batteries are depleted, **r** increases. If **r** becomes a significant fraction of the load resistance, then the current is significantly reduced, as the following example illustrates.

Example 5.2.1: Analyzing a Circuit with a Battery and a Load

A given battery has a 12.00-V emf and an internal resistance of $0.100\,\Omega$ (a) Calculate its terminal voltage when connected to a $10.00\,\Omega$ load. (b) What is the terminal voltage when connected to a $0.500\,\Omega$ load? (c) What power does the $0.500\,\Omega$ load dissipate? (d) If the internal resistance grows to $0.500\,\Omega$, find the current, terminal voltage, and power dissipated by a $0.500\,\Omega$ load.

Strategy

The analysis above gave an expression for current when internal resistance is taken into account. Once the current is found, the terminal voltage can be calculated by using the equation $V_{terminal} = \epsilon - Ir$. Once current is found, we can also find the power dissipated by the resistor.

Solution

1. Entering the given values for the emf, load resistance, and internal resistance into the expression above yields

$$I = \frac{\epsilon}{R+r} = \frac{12.00\,V}{10.10\,\Omega} = 1.188\,A.$$

   Enter the known values into the equation $V_{terminal} = \epsilon - Ir$ to get the terminal voltage:

$$V_{terminal} = \epsilon - Ir = 12.00\,V - (1.188\,A)(0.100\,\Omega) = 11.90\,V.$$

   The terminal voltage here is only slightly lower than the emf, implying that the current drawn by this light load is not significant.

2. Similarly, with $R_{load} = 0.500\,\Omega$, the current is

$$I = \frac{\epsilon}{R+r} = \frac{12.00\,V}{0.600\,\Omega} = 20.00\,A.$$

   The terminal voltage is now

$$V_{terminal} = \epsilon - Ir = 12.00\,V - (20.00\,A)(0.100\,\Omega) = 10.00\,V.$$

   The terminal voltage exhibits a more significant reduction compared with emf, implying $0.500\,\Omega$ is a heavy load for this battery. A “heavy load” signifies a larger draw of current from the source but not a larger resistance.

3. The power dissipated by the $0.500\,\Omega$ load can be found using the formula $P = I^2R$. Entering the known values gives

$$P = I^2R = (20.0\,A)^2(0.500\,\Omega) = 2.00 \times 10^2\,W.$$

Note that this power can also be obtained using the expression $\frac{V^2}{R}$ or IV, where **V** is the terminal voltage (10.0 V in this case).

4. Here, the internal resistance has increased, perhaps due to the depletion of the battery, to the point where it is as great as the load resistance. As before, we first find the current by entering the known values into the expression, yielding

$$I = \frac{\epsilon}{R+r} = \frac{12.00\,V}{1.00\,\Omega} = 12.00\,A.$$

Now the terminal voltage is

$$V_{terminal} = \epsilon - Ir = 12.00\,V - (12.00\,A)(0.500\,\Omega) = 6.00\,V,$$

and the power dissipated by the load is

$$P = I^2R = (12.00\,A)^2(0.500\,\Omega) = 72.00\,W.$$

We see that the increased internal resistance has significantly decreased the terminal voltage, current, and power delivered to a load.

Significance

The internal resistance of a battery can increase for many reasons. For example, the internal resistance of a rechargeable battery increases as the number of times the battery is recharged increases. The increased internal resistance may have two effects on the battery. First, the terminal voltage will decrease. Second, the battery may overheat due to the increased power dissipated by the internal resistance.

? Exercise 5.2.1

If you place a wire directly across the two terminal of a battery, effectively shorting out the terminals, the battery will begin to get hot. Why do you suppose this happens?

Solution

If a wire is connected across the terminals, the load resistance is close to zero, or at least considerably less than the internal resistance of the battery. Since the internal resistance is small, the current through the circuit will be large, $I = \frac{\epsilon}{R+r} = \frac{\epsilon}{0+r} = \frac{\epsilon}{r}$. The large current causes a high power to be dissipated by the internal resistance $(P = I^2r)$. The power is dissipated as heat.

Battery Testers

Battery testers, such as those in Figure 5.2.8, use small load resistors to intentionally draw current to determine whether the terminal potential drops below an acceptable level. Although it is difficult to measure the internal resistance of a battery, battery testers can provide a measurement of the internal resistance of the battery. If internal resistance is high, the battery is weak, as evidenced by its low terminal voltage.

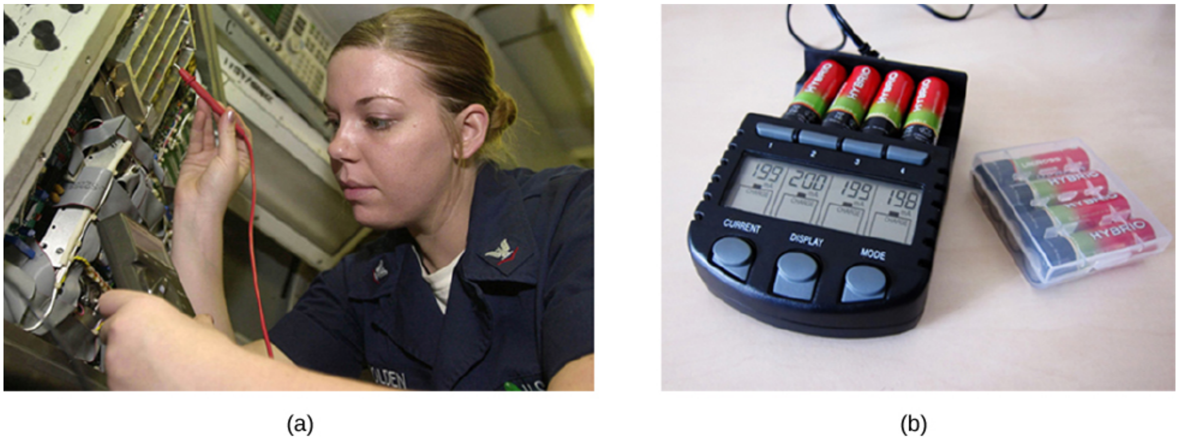


Figure 5.2.8: Battery testers measure terminal voltage under a load to determine the condition of a battery. (a) A US Navy electronics technician uses a battery tester to test large batteries aboard the aircraft carrier USS **Nimitz**. The battery tester she uses has a small resistance that can dissipate large amounts of power. (b) The small device shown is used on small batteries and has a digital display to indicate the acceptability of the terminal voltage. (credit a: modification of work by Jason A. Johnston; credit b: modification of work by Keith Williamson)

Some batteries can be recharged by passing a current through them in the direction opposite to the current they supply to an appliance. This is done routinely in cars and in batteries for small electrical appliances and electronic devices (Figure 5.2.9). The voltage output of the battery charger must be greater than the emf of the battery to reverse the current through it. This causes the terminal voltage of the battery to be greater than the emf, since $V = \epsilon - Ir$ and **I** is now negative.

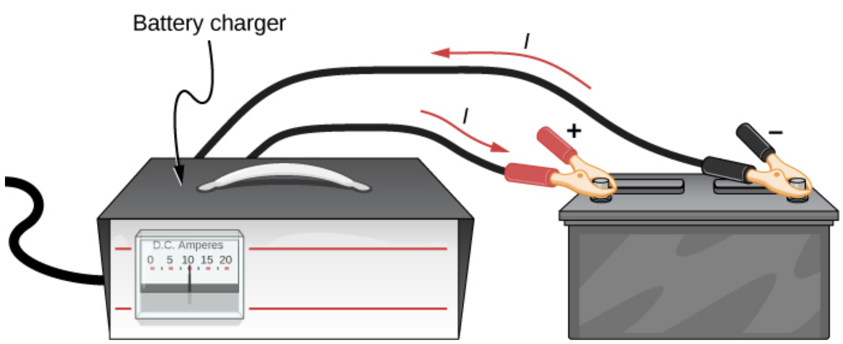


Figure 5.2.9: A car battery charger reverses the normal direction of current through a battery, reversing its chemical reaction and replenishing its chemical potential.

It is important to understand the consequences of the internal resistance of emf sources, such as batteries and solar cells, but often, the analysis of circuits is done with the terminal voltage of the battery, as we have done in the previous sections. The terminal voltage is referred to as simply as **V**, dropping the subscript "terminal." This is because the internal resistance of the battery is difficult to measure directly and can change over time.

5.3: Resistors in Series and Parallel

Learning Objectives

By the end of the section, you will be able to:

- Define the term equivalent resistance
- Calculate the equivalent resistance of resistors connected in series
- Calculate the equivalent resistance of resistors connected in parallel

In Current and Resistance, we described the term 'resistance' and explained the basic design of a resistor. Basically, a resistor limits the flow of charge in a circuit and is an ohmic device where $V = IR$. Most circuits have more than one resistor. If several resistors are connected together and connected to a battery, the current supplied by the battery depends on the **equivalent resistance** of the circuit.

The equivalent resistance of a combination of resistors depends on both their individual values and how they are connected. The simplest combinations of resistors are series and parallel connections (Figure $5.3.1$). In a **series circuit**, the output current of the first resistor flows into the input of the second resistor; therefore, the current is the same in each resistor. In a **parallel circuit**, all of the resistor leads on one side of the resistors are connected together and all the leads on the other side are connected together. In the case of a parallel configuration, each resistor has the same potential drop across it, and the currents through each resistor may be different, depending on the resistor. The sum of the individual currents equals the current that flows into the parallel connections.

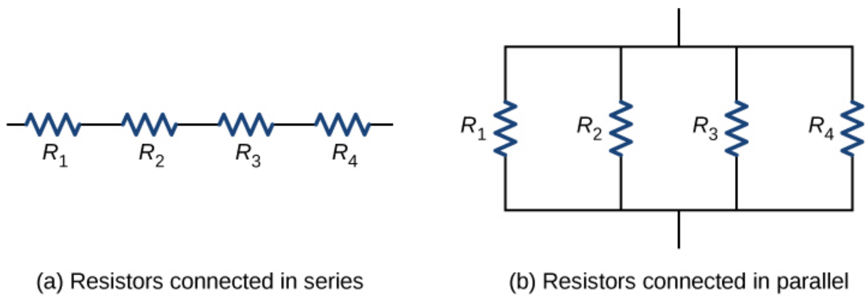


Figure $5.3.1$: (a) For a series connection of resistors, the current is the same in each resistor. (b) For a parallel connection of resistors, the voltage is the same across each resistor.

Resistors in Series

Resistors are said to be in series whenever the current flows through the resistors sequentially. Consider Figure $5.3.2$, which shows three resistors in series with an applied voltage equal to V_{ab}. Since there is only one path for the charges to flow through, the current is the same through each resistor. The equivalent resistance of a set of resistors in a series connection is equal to the algebraic sum of the individual resistances.

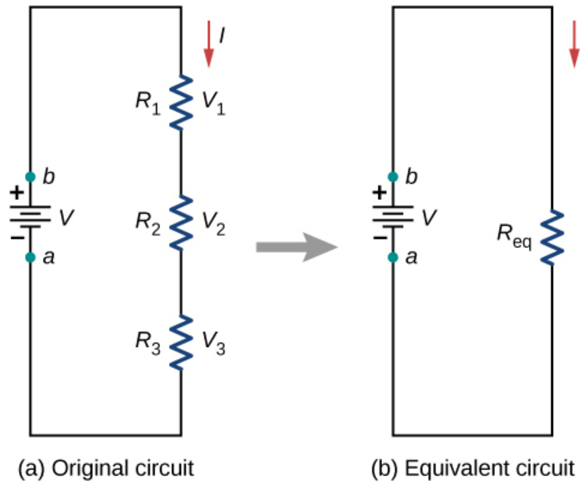


Figure 5.3.2: (a) Three resistors connected in series to a voltage source. (b) The original circuit is reduced to an equivalent resistance and a voltage source.

In Figure 5.3.2, the current coming from the voltage source flows through each resistor, so the current through each resistor is the same. The current through the circuit depends on the voltage supplied by the voltage source and the resistance of the resistors. For each resistor, a potential drop occurs that is equal to the loss of electric potential energy as a current travels through each resistor. According to Ohm's law, the potential drop V across a resistor when a current flows through it is calculated using the equation $V = IR$, where I is the current in amps (A) and R is the resistance in ohms (Ω). Since energy is conserved, and the voltage is equal to the potential energy per charge, the sum of the voltage applied to the circuit by the source and the potential drops across the individual resistors around a loop should be equal to zero:

$$\sum_{i=1}^{N} V_i = 0.$$

This equation is often referred to as Kirchhoff's loop law, which we will look at in more detail later in this chapter. For Figure 5.3.2, the sum of the potential drop of each resistor and the voltage supplied by the voltage source should equal zero:

$$\begin{aligned} V - V_1 - V_2 - V_3 &= 0, \\ V &= V_1 + V_2 + V_3, \\ &= IR_1 + IR_2 + IR_3, \end{aligned}$$

Solving for I

$$\begin{aligned} I &= \frac{V}{R_1 + R_2 + R_3} \\ &= \frac{V}{R_S}. \end{aligned}$$

Since the current through each component is the same, the equality can be simplified to an equivalent resistance (R_S), which is just the sum of the resistances of the individual resistors.

✔ Equivalent Resistance in Series Circuits

Any number of resistors can be connected in series. If N resistors are connected in series, the **equivalent resistance** is

$$R_S = R_1 + R_2 + R_3 + \ldots + R_{N-1} + R_N = \sum_{i=1}^{N} R_i. \tag{5.3.1}$$

One result of components connected in a series circuit is that if something happens to one component, it affects all the other components. For example, if several lamps are connected in series and one bulb burns out, all the other lamps go dark.

Example 5.3.1: Equivalent Resistance, Current, and Power in a Series Circuit

A battery with a terminal voltage of 9 V is connected to a circuit consisting of four $20\,\Omega$ and one $10\,\Omega$ resistors all in series (Figure 5.3.3). Assume the battery has negligible internal resistance.

a. Calculate the equivalent resistance of the circuit.
b. Calculate the current through each resistor.
c. Calculate the potential drop across each resistor.
d. Determine the total power dissipated by the resistors and the power supplied by the battery.

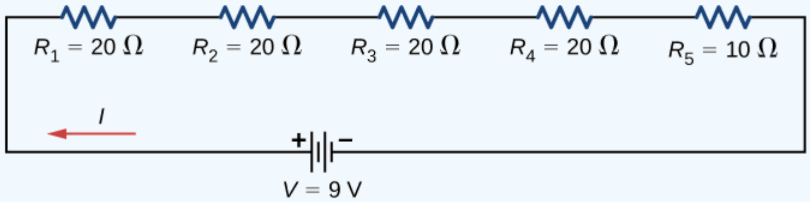


Figure 5.3.3: A simple series circuit with five resistors.

Strategy

In a series circuit, the equivalent resistance is the algebraic sum of the resistances. The current through the circuit can be found from Ohm's law and is equal to the voltage divided by the equivalent resistance. The potential drop across each resistor can be found using Ohm's law. The power dissipated by each resistor can be found using $P = I^2R$, and the total power dissipated by the resistors is equal to the sum of the power dissipated by each resistor. The power supplied by the battery can be found using $P = I\epsilon$.

Solution

1. The equivalent resistance is the algebraic sum of the resistances (Equation 5.3.1):

$$\begin{aligned} R_S &= R_1 + R_2 + R_3 + R_4 + R_5 \\ &= 20\,\Omega + 20\,\Omega + 20\,\Omega + 20\,\Omega + 10\,\Omega = 90\,\Omega. \end{aligned}$$

2. The current through the circuit is the same for each resistor in a series circuit and is equal to the applied voltage divided by the equivalent resistance:

$$I = \frac{V}{R_S} = \frac{9\,V}{90\,\Omega} = 0.1\,A.$$

Note that the sum of the potential drops across each resistor is equal to the voltage supplied by the battery.

3. The power dissipated by a resistor is equal to $P = I^2R$, and the power supplied by the battery is equal to $P = I\epsilon$.

$$\begin{aligned} P_1 = P_2 = P_3 = P_4 &= (0.1\,A)^2(20\,\Omega) = 0.2\,W, \\ P_5 &= (0.1\,A)^2(10\,\Omega) = 0.1\,W, \\ P_{dissipated} &= 0.2\,W + 0.2\,W + 0.2\,W + 0.2\,W + 0.1\,W = 0.9\,W, \\ P_{source} &= I\epsilon = (0.1\,A)(9\,V) = 0.9\,W. \end{aligned}$$

Significance

There are several reasons why we would use multiple resistors instead of just one resistor with a resistance equal to the equivalent resistance of the circuit. Perhaps a resistor of the required size is not available, or we need to dissipate the heat generated, or we want to minimize the cost of resistors. Each resistor may cost a few cents to a few dollars, but when multiplied by thousands of units, the cost saving may be appreciable.

? Exercise 5.3.1

Some strings of miniature holiday lights are made to short out when a bulb burns out. The device that causes the short is called a shunt, which allows current to flow around the open circuit. A "short" is like putting a piece of wire across the component.

The bulbs are usually grouped in series of nine bulbs. If too many bulbs burn out, the shunts eventually open. What causes this?

Answer

The equivalent resistance of nine bulbs connected in series is 9**R**. The current is $I = V/9\,R$. If one bulb burns out, the equivalent resistance is 8**R**, and the voltage does not change, but the current increases $(I = V/8\,R$. As more bulbs burn out, the current becomes even higher. Eventually, the current becomes too high, burning out the shunt.

Let's briefly summarize the major features of resistors in series:

1. Series resistances add together to get the equivalent resistance (Equation 5.3.1):

$$R_S = R_1 + R_2 + R_3 + \ldots + R_{N-1} + R_N = \sum_{i=1}^{N} R_i.$$

2. The same current flows through each resistor in series.
3. Individual resistors in series do not get the total source voltage, but divide it. The total potential drop across a series configuration of resistors is equal to the sum of the potential drops across each resistor.

Resistors in Parallel

Figure 5.3.4 shows resistors in parallel, wired to a voltage source. Resistors are in parallel when one end of all the resistors are connected by a continuous wire of negligible resistance and the other end of all the resistors are also connected to one another through a continuous wire of negligible resistance. The potential drop across each resistor is the same. Current through each resistor can be found using Ohm's law $I = V/R$, where the voltage is constant across each resistor. For example, an automobile's headlights, radio, and other systems are wired in parallel, so that each subsystem utilizes the full voltage of the source and can operate completely independently. The same is true of the wiring in your house or any building.

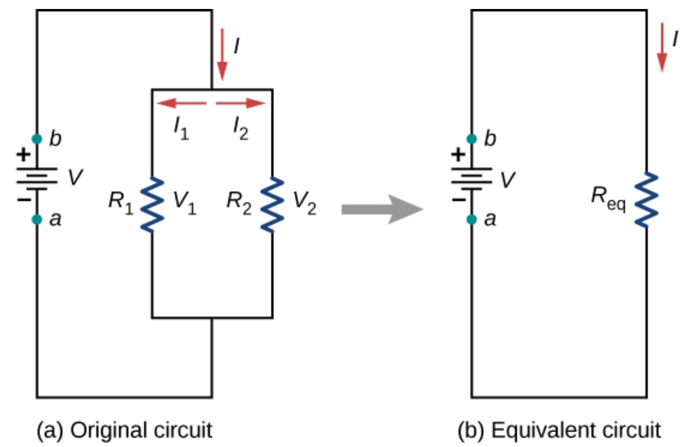


Figure 5.3.4: Two resistors connected in parallel to a voltage source. (b) The original circuit is reduced to an equivalent resistance and a voltage source.

The current flowing from the voltage source in Figure 5.3.4 depends on the voltage supplied by the voltage source and the equivalent resistance of the circuit. In this case, the current flows from the voltage source and enters a junction, or node, where the circuit splits flowing through resistors R_1 and R_2. As the charges flow from the battery, some go through resistor R_1 and some flow through resistor R_2. The sum of the currents flowing into a junction must be equal to the sum of the currents flowing out of the junction:

$$\sum I_{in} = \sum I_{out}.$$

This equation is referred to as **Kirchhoff's junction rule** and will be discussed in detail in the next section. In Figure 5.3.4, the junction rule gives $I = I_1 + I_2$. There are two loops in this circuit, which leads to the equations $V = I_1 R_1$ and $I_1 R_1 = I_2 R_2$. Note the voltage across the resistors in parallel are the same ($V = V_1 = V_2$) and the current is additive:

$$\begin{aligned} I &= I_1 + I_2 \\ &= \frac{V_1}{R_1} + \frac{V_2}{R_2} \\ &= \frac{V}{R_1} + \frac{V}{R_2} \\ &= V\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{V}{R_P} \end{aligned}$$

Solving for the R_P

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

Equivalent Resistance in Parallel Circuits

Generalizing to any number of N resistors, the equivalent resistance R_P of a parallel connection is related to the individual resistances by

$$R_P = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots + \frac{1}{R_{N-1}} + \frac{1}{R_N}\right)^{-1} = \left(\sum_{i=1}^{N} \frac{1}{R_i}\right)^{-1}. \tag{5.3.2}$$

This relationship results in an equivalent resistance R_P that is less than the smallest of the individual resistances. When resistors are connected in parallel, more current flows from the source than would flow for any of them individually, so the total resistance is lower.

Example 5.3.2: Analysis of a parallel circuit

Three resistors $R_1 = 1.00\,\Omega$, $R_2 = 2.00\,\Omega$, and $R_3 = 2.00\,\Omega$, are connected in parallel. The parallel connection is attached to a $V = 3.00\,V$ voltage source.

a. What is the equivalent resistance?
b. Find the current supplied by the source to the parallel circuit.
c. Calculate the currents in each resistor and show that these add together to equal the current output of the source.
d. Calculate the power dissipated by each resistor.
e. Find the power output of the source and show that it equals the total power dissipated by the resistors.

Strategy

(a) The total resistance for a parallel combination of resistors is found using Equation 5.3.2. (Note that in these calculations, each intermediate answer is shown with an extra digit.)

(b) The current supplied by the source can be found from Ohm's law, substituting R_P for the total resistance $I = \frac{V}{R_P}$.

(c) The individual currents are easily calculated from Ohm's law $\left(I_i = \frac{V_i}{R_i}\right)$, since each resistor gets the full voltage. The total current is the sum of the individual currents:

$$I = \sum_i I_i.$$

(d) The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three are known. Let us use $P_i = V^2/R_i$, since each resistor gets full voltage.

(e) The total power can also be calculated in several ways, use $P = IV$.

Solution

1. The total resistance for a parallel combination of resistors is found using Equation 5.3.2. Entering known values gives

$$R_P = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right)^{-1} = \left(\frac{1}{1.00\,\Omega} + \frac{1}{2.00\,\Omega} + \frac{1}{2.00\,\Omega}\right)^{-1} = 0.50\,\Omega.$$

The total resistance with the correct number of significant digits is $R_{eq} = 0.50\,\Omega$. As predicted, R_P is less than the smallest individual resistance.

2. The total current can be found from Ohm's law, substituting R_P for the total resistance. This gives

$$I = \frac{V}{R_P} = \frac{3.00\,V}{0.50\,\Omega} = 6.00\,A.$$

Current **I** for each device is much larger than for the same devices connected in series (see the previous example). A circuit with parallel connections has a smaller total resistance than the resistors connected in series.

3. The individual currents are easily calculated from Ohm's law, since each resistor gets the full voltage. Thus,

$$I_1 = \frac{V}{R_1} = \frac{3.00\,V}{1.00\,\Omega} = 3.00\,A.$$

Similarly,

$$I_2 = \frac{V}{R_2} = \frac{3.00\,V}{2.00\,\Omega} = 1.50\,A$$

and

$$I_3 = \frac{V}{R_3} = \frac{3.00\,V}{2.00\,\Omega} = 1.50\,A.$$

The total current is the sum of the individual currents:

$$I_1 + I_2 + I_3 = 6.00\,A.$$

4. The power dissipated by each resistor can be found using any of the equations relating power to current, voltage, and resistance, since all three are known. Let us use $P = V^2/R$, since each resistor gets full voltage. Thus,

$$P_1 = \frac{V^2}{R_1} = \frac{(3.00\,V)^2}{1.00\,\Omega} = 9.00\,W.$$

Similarly,

$$P_2 = \frac{V^2}{R_2} = \frac{(3.00\,V)^2}{2.00\,\Omega} = 4.50\,W.$$

and

$$P_3 = \frac{V^2}{R_3} = \frac{(3.00\,V)^2}{2.00\,\Omega} = 4.50\,W.$$

5. The total power can also be calculated in several ways. Choosing $P = IV$ and entering the total current yields

$$P = IV = (6.00\,A)(3.00\,V) = 18.00\,W.$$

Significance

Total power dissipated by the resistors is also 18.00 W:

$$P_1 + P_2 + P_3 = 9.00\,W + 4.50\,W + 4.50\,W = 18.00\,W.$$

Notice that the total power dissipated by the resistors equals the power supplied by the source.

? Exercise 5.3.2A

Consider the same potential difference $(V = 3.00\,V)$ applied to the same three resistors connected in series. Would the equivalent resistance of the series circuit be higher, lower, or equal to the three resistor in parallel? Would the current through the series circuit be higher, lower, or equal to the current provided by the same voltage applied to the parallel circuit? How would the power dissipated by the resistor in series compare to the power dissipated by the resistors in parallel?

Solution

The equivalent of the series circuit would be $R_{eq} = 1.00\,\Omega + 2.00\,\Omega + 2.00\,\Omega = 5.00\,\Omega$, which is higher than the equivalent resistance of the parallel circuit $R_{eq} = 0.50\,\Omega$. The equivalent resistor of any number of resistors is always higher than the equivalent resistance of the same resistors connected in parallel. The current through for the series circuit would be $I = \frac{3.00\,V}{5.00\,\Omega} = 0.60\,A$, which is lower than the sum of the currents through each resistor in the parallel circuit, $I = 6.00\,A$. This is not surprising since the equivalent resistance of the series circuit is higher. The current through a series connection of any number of resistors will always be lower than the current into a parallel connection of the same resistors, since the equivalent resistance of the series circuit will be higher than the parallel circuit. The power dissipated by the resistors in series would be $P = 1.800\,W$, which is lower than the power dissipated in the parallel circuit $P = 18.00\,W$.

Exercise $5.3.2B$

How would you use a river and two waterfalls to model a parallel configuration of two resistors? How does this analogy break down?

Solution

A river, flowing horizontally at a constant rate, splits in two and flows over two waterfalls. The water molecules are analogous to the electrons in the parallel circuits. The number of water molecules that flow in the river and falls must be equal to the number of molecules that flow over each waterfall, just like sum of the current through each resistor must be equal to the current flowing into the parallel circuit. The water molecules in the river have energy due to their motion and height. The potential energy of the water molecules in the river is constant due to their equal heights. This is analogous to the constant change in voltage across a parallel circuit. Voltage is the potential energy across each resistor.

The analogy quickly breaks down when considering the energy. In the waterfall, the potential energy is converted into kinetic energy of the water molecules. In the case of electrons flowing through a resistor, the potential drop is converted into heat and light, not into the kinetic energy of the electrons.

Let us summarize the major features of resistors in parallel:

1. Equivalent resistance is found from Equation $5.3.2$ and is smaller than any individual resistance in the combination.
2. The potential drop across each resistor in parallel is the same.
3. Parallel resistors do not each get the total current; they divide it. The current entering a parallel combination of resistors is equal to the sum of the current through each resistor in parallel.

In this chapter, we introduced the equivalent resistance of resistors connect in series and resistors connected in parallel. You may recall from the Section on Capacitance, we introduced the equivalent capacitance of capacitors connected in series and parallel. Circuits often contain both capacitors and resistors. Table $5.3.1$ summarizes the equations used for the equivalent resistance and equivalent capacitance for series and parallel connections.

Table $5.3.1$: Summary for Equivalent Resistance and Capacitance in Series and Parallel Combinations

	Series combination	Parallel combination
Equivalent capacitance	$\frac{1}{C_S} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$	$C_P = C_1 + C_2 + C_3 + \ldots$
Equivalent resistance	$R_S = R_1 + R_2 + R_3 + \ldots = \sum_{i=1}^{N} R_i$	$\frac{1}{R_P} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$

Combinations of Series and Parallel

More complex connections of resistors are often just combinations of series and parallel connections. Such combinations are common, especially when wire resistance is considered. In that case, wire resistance is in series with other resistances that are in parallel.

Combinations of series and parallel can be reduced to a single equivalent resistance using the technique illustrated in Figure $5.3.5$. Various parts can be identified as either series or parallel connections, reduced to their equivalent resistances, and then further reduced until a single equivalent resistance is left. The process is more time consuming than difficult. Here, we note the equivalent resistance as R_{eq}.

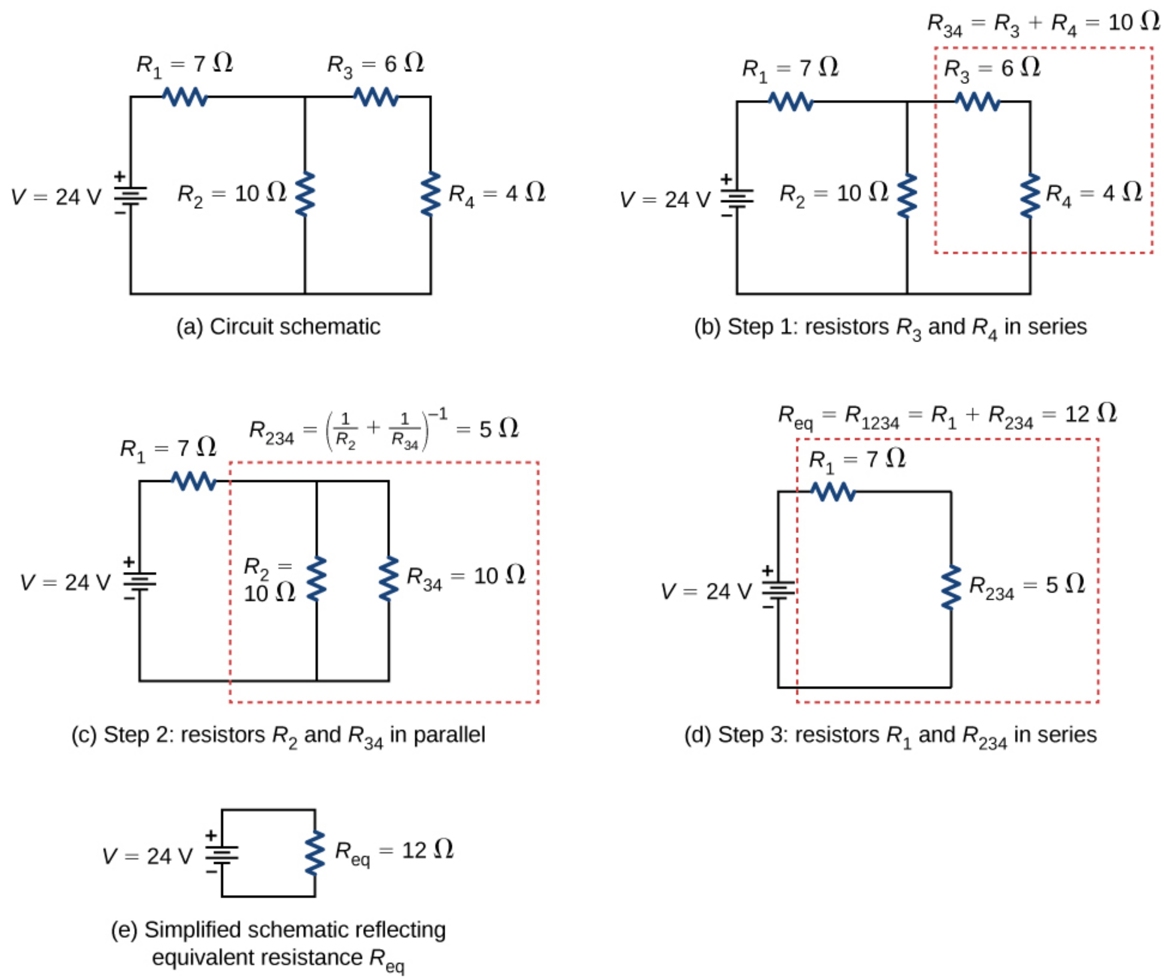


Figure $5.3.5$: (a) The original circuit of four resistors. (b) Step 1: The resistors R_3 and R_4 are in series and the equivalent resistance is $R_{34} = 10\,\Omega$ (c) Step 2: The reduced circuit shows resistors R_2 and R_{34} are in parallel, with an equivalent resistance of $R_{234} = 5\,\Omega$. (d) Step 3: The reduced circuit shows that R_1 and R_{234} are in series with an equivalent resistance of $R_{1234} = 12\,\Omega$ which is the equivalent resistance R_{eq}. (e) The reduced circuit with a voltage source of $V = 24\,V$ with an equivalent resistance of $R_{eq} = 12\,\Omega$. This results in a current of $I = 2\,A$ from the voltage source.

Notice that resistors R_3 and R_4 are in series. They can be combined into a single equivalent resistance. One method of keeping track of the process is to include the resistors as subscripts. Here the equivalent resistance of R_3 and R_4 is

$$R_{34} = R_3 + R_4 = 6\,\Omega + 4\,\Omega = 10\,\Omega.$$

The circuit now reduces to three resistors, shown in Figure $5.3.5c$. Redrawing, we now see that resistors R_2 and R_{34} constitute a parallel circuit. Those two resistors can be reduced to an equivalent resistance:

$$R_{234} = \left(\frac{1}{R_2} + \frac{1}{R_{34}}\right)^{-1} = \left(\frac{1}{10\,\Omega} + \frac{1}{10\,\Omega}\right)^{-1} = 5\,\Omega.$$

This step of the process reduces the circuit to two resistors, shown in in Figure $5.3.5d$. Here, the circuit reduces to two resistors, which in this case are in series. These two resistors can be reduced to an equivalent resistance, which is the equivalent resistance of the circuit:

$$R_{eq} = R_{1234} = R_1 + R_{234} = 7\,\Omega + 5\Omega = 12\,\Omega.$$

The main goal of this circuit analysis is reached, and the circuit is now reduced to a single resistor and single voltage source.

Now we can analyze the circuit. The current provided by the voltage source is $I = \frac{V}{R_{eq}} = \frac{24\,V}{12\,\Omega} = 2\,A$. This current runs through resistor R_1 and is designated as I_1. The potential drop across R_1 can be found using Ohm's law:

$$V_1 = I_1 R_1 = (2\,A)(7\,\Omega) = 14\,V.$$

Looking at Figure $5.3.5c$, this leaves $24\,V - 14\,V = 10\,V$ to be dropped across the parallel combination of R_2 and R_{34}. The current through R_2 can be found using Ohm's law:

$$I_2 = \frac{V_2}{R_2} = \frac{10\,V}{10\,\Omega} = 1\,A.$$

The resistors R_3 and R_4 are in series so the currents I_3 and I_4 are equal to

$$I_3 = I_4 = I - I_2 = 2\,A - 1\,A = 1\,A.$$

Using Ohm's law, we can find the potential drop across the last two resistors. The potential drops are $V_3 = I_3 R_3 = 6\,V$ and $V_4 = I_4 R_4 = 4\,V$. The final analysis is to look at the power supplied by the voltage source and the power dissipated by the resistors. The power dissipated by the resistors is

$$\begin{aligned} P_1 &= I_1^2 R_1 = (2\,A)^2 (7\,\Omega) = 28\,W, \\ P_2 &= I_2^2 R_2 = (1\,A)^2 (10\,\Omega) = 10\,W, \\ P_3 &= I_3^2 R_3 = (1\,A)^2 (6\,\Omega) = 6\,W, \\ P_4 &= I_4^2 R_4 = (1\,A)^2 (4\,\Omega) = 4\,W, \\ P_{dissipated} &= P_1 + P_2 + P_3 + P_4 = 48\,W. \end{aligned}$$

The total energy is constant in any process. Therefore, the power supplied by the voltage source is

$$\begin{aligned} P_s &= IV \\ &= (2\,A)(24\,V) = 48\,W \end{aligned}$$

Analyzing the power supplied to the circuit and the power dissipated by the resistors is a good check for the validity of the analysis; they should be equal.

✓ Example $5.3.3$: Combining Series and parallel circuits

Figure $5.3.6$ shows resistors wired in a combination of series and parallel. We can consider R_1 to be the resistance of wires leading to R_2 and R_3.

a. Find the equivalent resistance of the circuit.
b. What is the potential drop V_1 across resistor R_1?
c. Find the current I_2 through resistor R_2.
d. What power is dissipated by R_2?

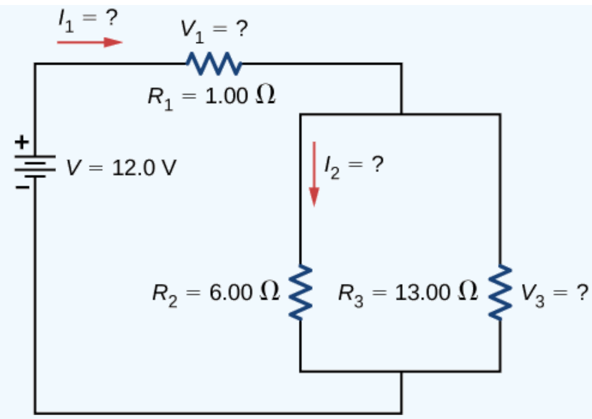


Figure $5.3.6$: These three resistors are connected to a voltage source so that R_2 and R_3 are in parallel with one another and that combination is in series with R_1.

Strategy

(a) To find the equivalent resistance, first find the equivalent resistance of the parallel connection of R_2 and R_3. Then use this result to find the equivalent resistance of the series connection with R_1.

(b) The current through R_1 can be found using Ohm's law and the voltage applied. The current through R_1 is equal to the current from the battery. The potential drop V_1 across the resistor R_1 (which represents the resistance in the connecting wires) can be found using Ohm's law.

(c) The current through R_2 can be found using Ohm's law $I_2 = \frac{V_2}{R_2}$. The voltage across R_2 can be found using $V_2 = V - V_1$.

(d) Using Ohm's law $(V_2 = I_2 R_2)$, the power dissipated by the resistor can also be found using $P_2 = I_2^2 R_2 = \frac{V_2^2}{R_2}$.

Solution

1. To find the equivalent resistance of the circuit, notice that the parallel connection of R_2 and R_3 is in series with R_1, so the equivalent resistance is

$$R_{eq} = R_1 + \left(\frac{1}{R_2} + \frac{1}{R_3}\right)^{-1} = 1.00\,\Omega + \left(\frac{1}{6.00\,\Omega} + \frac{1}{13.00\,\Omega}\right)^{-1} = 5.10\,\Omega.$$

   The total resistance of this combination is intermediate between the pure series and pure parallel values ($20.0\,\Omega$ and $0.804\,\Omega$, respectively).
2. The current through R_1 is equal to the current supplied by the battery:

$$I_1 = I = \frac{V}{R_{eq}} = \frac{12.0\,V}{5.10\,\Omega} = 2.35\,A.$$

   The voltage across R_1 is

$$V_1 = I_1 R_1 = (2.35\,A)(1\,\Omega) = 2.35\,V.$$

   The voltage applied to R_2 and R_3 is less than the voltage supplied by the battery by an amount V_1. When wire resistance is large, it can significantly affect the operation of the devices represented by R_2 and R_3.
3. To find the current through R_2, we must first find the voltage applied to it. The voltage across the two resistors in parallel is the same:

$$V_2 = V_3 = V - V_1 = 12.0\,V - 2.35\,V = 9.65\,V.$$

   Now we can find the current I_2 through resistance R_2 using Ohm's law:

$$I_2 = \frac{V_2}{R_2} = \frac{9.65\,V}{6.00\,\Omega} = 1.61\,A.$$

   The current is less than the 2.00 A that flowed through R_2 when it was connected in parallel to the battery in the previous parallel circuit example.

4. The power dissipated by R_2 is given by

$$P_2 = I_2^2 R_2 = (1.61\,A)^2(6.00\,\Omega) = 15.5\,W.$$

Significance

The analysis of complex circuits can often be simplified by reducing the circuit to a voltage source and an equivalent resistance. Even if the entire circuit cannot be reduced to a single voltage source and a single equivalent resistance, portions of the circuit may be reduced, greatly simplifying the analysis.

Exercise $5.3.3$

Consider the electrical circuits in your home. Give at least two examples of circuits that must use a combination of series and parallel circuits to operate efficiently.

Solution

All the overhead lighting circuits are in parallel and connected to the main supply line, so when one bulb burns out, all the overhead lighting does not go dark. Each overhead light will have at least one switch in series with the light, so you can turn it on and off.

A refrigerator has a compressor and a light that goes on when the door opens. There is usually only one cord for the refrigerator to plug into the wall. The circuit containing the compressor and the circuit containing the lighting circuit are in parallel, but there is a switch in series with the light. A thermostat controls a switch that is in series with the compressor to control the temperature of the refrigerator.

Practical Implications

One implication of this last example is that resistance in wires reduces the current and power delivered to a resistor. If wire resistance is relatively large, as in a worn (or a very long) extension cord, then this loss can be significant. If a large current is drawn, the **IR** drop in the wires can also be significant and may become apparent from the heat generated in the cord.

For example, when you are rummaging in the **refrigerator** and the motor comes on, the refrigerator light dims momentarily. Similarly, you can see the passenger compartment light dim when you start the engine of your car (although this may be due to resistance inside the battery itself).

What is happening in these high-current situations is illustrated in Figure $5.3.7$. The device represented by R_3 has a very low resistance, so when it is switched on, a large current flows. This increased current causes a larger **IR** drop in the wires represented by R_1, reducing the voltage across the light bulb (which is R_2), which then dims noticeably.

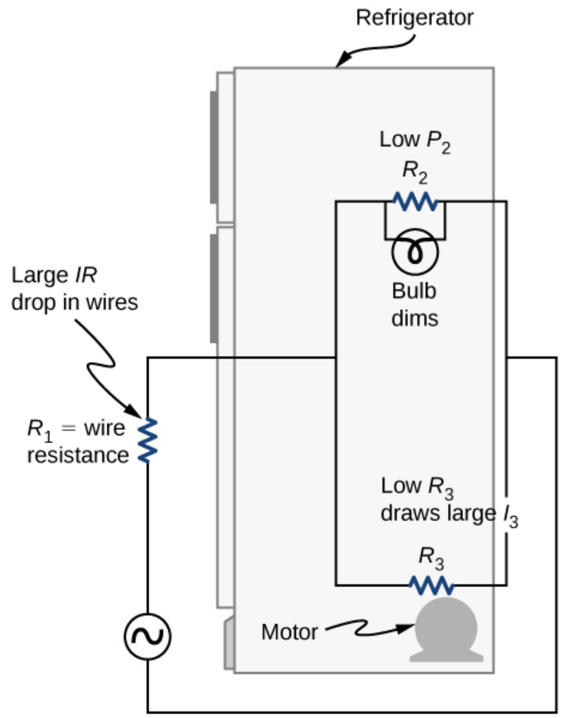


Figure 5.3.7: Why do lights dim when a large appliance is switched on? The answer is that the large current the appliance motor draws causes a significant **IR** drop in the wires and reduces the voltage across the light.

✓ Problem-Solving Strategy: Series and Parallel Resistors

1. Draw a clear circuit diagram, labeling all resistors and voltage sources. This step includes a list of the known values for the problem, since they are labeled in your circuit diagram.
2. Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.
3. Determine whether resistors are in series, parallel, or a combination of both series and parallel. Examine the circuit diagram to make this assessment. Resistors are in series if the same current must pass sequentially through them.
4. Use the appropriate list of major features for series or parallel connections to solve for the unknowns. There is one list for series and another for parallel.
5. Check to see whether the answers are reasonable and consistent.

✓ Example 5.3.4: Combining Series and Parallel circuits

Two resistors connected in series (R_1, R_2) are connected to two resistors that are connected in parallel (R_3, R_4). The series-parallel combination is connected to a battery. Each resistor has a resistance of 10.00 Ohms. The wires connecting the resistors and battery have negligible resistance. A current of 2.00 Amps runs through resistor R_1. What is the voltage supplied by the voltage source?

Strategy

Use the steps in the preceding problem-solving strategy to find the solution for this example.

Solution

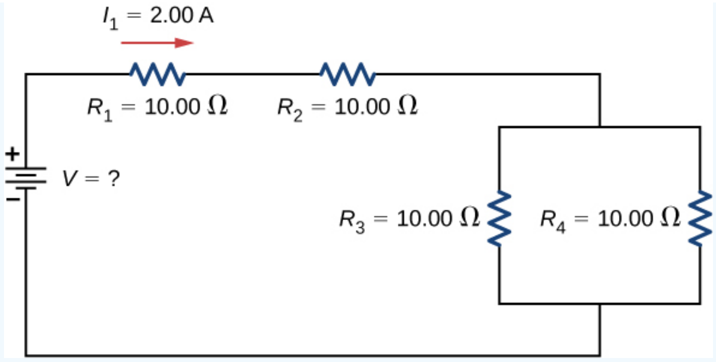


Figure $5.3.8$: To find the unknown voltage, we must first find the equivalent resistance of the circuit.

1. Draw a clear circuit diagram (Figure $5.3.8$).
2. The unknown is the voltage of the battery. In order to find the voltage supplied by the battery, the equivalent resistance must be found.
3. In this circuit, we already know that the resistors R_1 and R_2 are in series and the resistors R_3 and R_4 are in parallel. The equivalent resistance of the parallel configuration of the resistors R_3 and R_4 is in series with the series configuration of resistors R_1 and R_2.
4. The voltage supplied by the battery can be found by multiplying the current from the battery and the equivalent resistance of the circuit. The current from the battery is equal to the current through R_1 and is equal to 2.00 A. We need to find the equivalent resistance by reducing the circuit. To reduce the circuit, first consider the two resistors in parallel. The equivalent resistance is

$$R_{34} = \left(\frac{1}{10.00\,\Omega} + \frac{1}{10.00\,\Omega}\right)^{-1} = 5.00\,\Omega.$$

   This parallel combination is in series with the other two resistors, so the equivalent resistance of the circuit is $R_{eq} = R_1 + R_2 + R_{34} = (25.00\,\Omega$. The voltage supplied by the battery is therefore $V = IR_{eq} = 2.00\ A(25.00\,\Omega) = 50.00\ V.$
5. One way to check the consistency of your results is to calculate the power supplied by the battery and the power dissipated by the resistors. The power supplied by the battery is $P_{batt} = IV = 100.00\ W$.

Since they are in series, the current through R_2 equals the current through R_1. Since $R_3 = R_4$, the current through each will be 1.00 Amps. The power dissipated by the resistors is equal to the sum of the power dissipated by each resistor:

$$\begin{aligned} P &= I_1^2 R_1 + I_2^2 R_2 + I_3^2 R_3 + I_4^2 R_4 \\ &= 40.00\ W + 40.00\ W + 10.00\ W + 10.00\ W = 100.\ W. \end{aligned}$$

Since the power dissipated by the resistors equals the power supplied by the battery, our solution seems consistent.

Significance

If a problem has a combination of series and parallel, as in this example, it can be reduced in steps by using the preceding problem-solving strategy and by considering individual groups of series or parallel connections. When finding R_{eq} for a parallel connection, the reciprocal must be taken with care. In addition, units and numerical results must be reasonable. Equivalent series resistance should be greater, whereas equivalent parallel resistance should be smaller, for example. Power should be greater for the same devices in parallel compared with series, and so on.

CHAPTER OVERVIEW

6: Magnetic Fields

For the past few chapters, we have been studying electrostatic forces and fields, which are caused by electric charges at rest. These electric fields can move other free charges, such as producing a current in a circuit; however, the electrostatic forces and fields themselves come from other static charges. In this chapter, we see that when an electric charge moves, it generates other forces and fields. These additional forces and fields are what we commonly call magnetism.

6.1: Prelude to Magnetic Forces and Fields

For the past few chapters, we have been studying electrostatic forces and fields, which are caused by electric charges at rest. These electric fields can move other free charges, such as producing a current in a circuit; however, the electrostatic forces and fields themselves come from other static charges. In this chapter, we see that when an electric charge moves, it generates other forces and fields. These additional forces and fields are what we commonly call magnetism.

Figure 6.1.1: An industrial electromagnet is capable of lifting thousands of pounds of metallic waste. (credit: modification of work by “BedfordAl”/Flickr)

Before we examine the origins of magnetism, we first describe what it is and how magnetic fields behave. Once we are more familiar with magnetic effects, we can explain how they arise from the behavior of atoms and molecules, and how magnetism is related to electricity. The connection between electricity and magnetism is fascinating from a theoretical point of view, but it is also immensely practical, as shown by an industrial electromagnet that can lift thousands of pounds of metal.

6.2: Magnetism and Its Historical Discoveries

Learning Objectives

By the end of this section, you will be able to:

- Explain attraction and repulsion by magnets
- Describe the historical and contemporary applications of magnetism

Magnetism has been known since the time of the ancient Greeks, but it has always been a bit mysterious. You can see electricity in the flash of a lightning bolt, but when a compass needle points to magnetic north, you can't see any force causing it to rotate. People learned about magnetic properties gradually, over many years, before several physicists of the nineteenth century connected magnetism with electricity. In this section, we review the basic ideas of magnetism and describe how they fit into the picture of a magnetic field.

Brief History of Magnetism

Magnets are commonly found in everyday objects, such as toys, hangers, elevators, doorbells, and computer devices. Experimentation on these magnets shows that all magnets have two poles: One is labeled north (N) and the other is labeled south (S). Magnetic poles repel if they are alike (both N or both S), they attract if they are opposite (one N and the other S), and both poles of a magnet attract unmagnetized pieces of iron. An important point to note here is that you cannot isolate an individual magnetic pole. Every piece of a magnet, no matter how small, which contains a north pole must also contain a south pole.

Note

Visit this website for an interactive demonstration of magnetic north and south poles.

An example of a magnet is a **compass needle**. It is simply a thin bar magnet suspended at its center, so it is free to rotate in a horizontal plane. Earth itself also acts like a very large bar magnet, with its south-seeking pole near the geographic North Pole (Figure 6.2.1). The north pole of a compass is attracted toward Earth's geographic North Pole because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole. Confusion arises because the geographic term "North Pole" has come to be used (incorrectly) for the magnetic pole that is near the North Pole. Thus, "**north magnetic pole**" is actually a misnomer—it should be called the **south magnetic pole**. [Note that the orientation of Earth's magnetic field is not permanent but changes ("flips") after long time intervals. Eventually, Earth's north magnetic pole may be located near its geographic North Pole.]

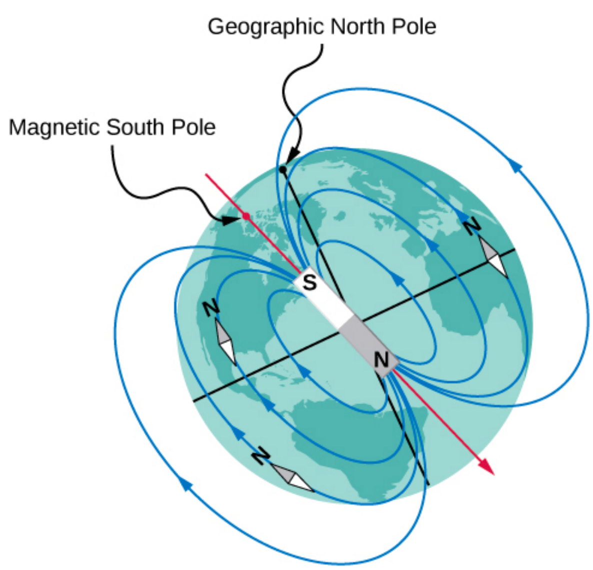


Figure 6.2.1: The north pole of a compass needle points toward the south pole of a magnet, which is how today's magnetic field is oriented from inside Earth. It also points toward Earth's geographic North Pole because the geographic North Pole is near the magnetic south pole.

Back in 1819, the Danish physicist Hans **Oersted** was performing a lecture demonstration for some students and noticed that a compass needle moved whenever current flowed in a nearby wire. Further investigation of this phenomenon convinced Oersted that an electric current could somehow cause a magnetic force. He reported this finding to an 1820 meeting of the French Academy of Science.

Soon after this report, Oersted's investigations were repeated and expanded upon by other scientists. Among those whose work was especially important were Jean-Baptiste **Biot** and Felix **Savart**, who investigated the forces exerted on magnets by currents; André Marie **Ampère**, who studied the forces exerted by one current on another; François **Arago**, who found that iron could be magnetized by a current; and Humphry **Davy**, who discovered that a magnet exerts a force on a wire carrying an electric current. Within 10 years of Oersted's discovery, Michael **Faraday** found that the relative motion of a magnet and a metallic wire induced current in the wire. This finding showed not only that a current has a magnetic effect, but that a magnet can generate electric current. You will see later that the names of Biot, Savart, Ampère, and Faraday are linked to some of the fundamental laws of electromagnetism.

The evidence from these various experiments led Ampère to propose that electric current is the source of all magnetic phenomena. To explain permanent magnets, he suggested that matter contains microscopic current loops that are somehow aligned when a material is magnetized. Today, we know that permanent magnets are actually created by the alignment of spinning electrons, a situation quite similar to that proposed by Ampère. This model of permanent magnets was developed by Ampère almost a century before the atomic nature of matter was understood. (For a full quantum mechanical treatment of magnetic spins, see Quantum Mechanics and Atomic Structure.)

Contemporary Applications of Magnetism

Today, magnetism plays many important roles in our lives. Physicists' understanding of magnetism has enabled the development of technologies that affect both individuals and society. The electronic tablet in your purse or backpack, for example, wouldn't have been possible without the applications of magnetism and electricity on a small scale (Figure $6.2.2$). Weak changes in a magnetic field in a thin film of iron and chromium were discovered to bring about much larger changes in resistance, called **giant magnetoresistance**. Information can then be recorded magnetically based on the direction in which the iron layer is magnetized. As a result of the discovery of giant magnetoresistance and its applications to digital storage, the 2007 Nobel Prize in Physics was awarded to Albert Fert from France and Peter Grunberg from Germany.

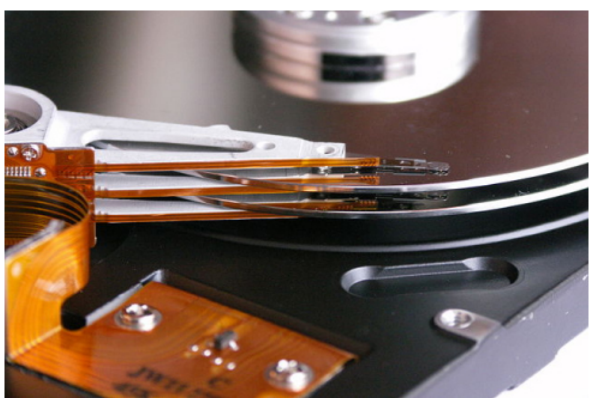


Figure $6.2.2$: Engineering technology like computer storage would not be possible without a deep understanding of magnetism. (credit: Klaus Eifert)

All electric motors—with uses as diverse as powering refrigerators, starting cars, and moving elevators—contain magnets. Generators, whether producing hydroelectric power or running bicycle lights, use magnetic fields. Recycling facilities employ magnets to separate iron from other refuse. Research into using magnetic containment of fusion as a future energy source has been continuing for several years. Magnetic resonance imaging (MRI) has become an important diagnostic tool in the field of medicine, and the use of magnetism to explore brain activity is a subject of contemporary research and development. The list of applications also includes computer hard drives, tape recording, detection of inhaled asbestos, and levitation of high-speed trains. Magnetism is involved in the structure of atomic energy levels, as well as the motion of cosmic rays and charged particles trapped in the Van Allen belts around Earth. Once again, we see that all these disparate phenomena are linked by a small number of underlying physical principles.

6.3: Magnetic Fields and Lines

Learning Objectives

By the end of this section, you will be able to:

- Define the magnetic field based on a moving charge experiencing a force
- Apply the right-hand rule to determine the direction of a magnetic force based on the motion of a charge in a magnetic field
- Sketch magnetic field lines to understand which way the magnetic field points and how strong it is in a region of space

We have outlined the properties of magnets, described how they behave, and listed some of the applications of magnetic properties. Even though there are no such things as isolated magnetic charges, we can still define the attraction and repulsion of magnets as based on a field. In this section, we define the magnetic field, determine its direction based on the right-hand rule, and discuss how to draw magnetic field lines.

Defining the Magnetic Field

A magnetic field is defined by the force that a charged particle experiences moving in this field, after we account for the gravitational and any additional electric forces possible on the charge. The magnitude of this force is proportional to the amount of charge **q**, the speed of the charged particle **v**, and the magnitude of the applied magnetic field. The direction of this force is perpendicular to both the direction of the moving charged particle and the direction of the applied magnetic field. Based on these observations, we define the magnetic field strength **B** based on the **magnetic force** $\vec{F}$ on a charge **q** moving at velocity as the cross product of the velocity and magnetic field, that is,

$$\vec{F} = q\vec{v} \times \vec{B}. \tag{6.3.1}$$

In fact, this is how we define the magnetic field $\vec{B}$ - in terms of the force on a charged particle moving in a magnetic field. The magnitude of the force is determined from the definition of the cross product as it relates to the magnitudes of each of the vectors. In other words, the magnitude of the force satisfies

$$F = qvB\sin\theta \tag{6.3.2}$$

where **θ** is the angle between the velocity and the magnetic field.

The SI unit for magnetic field strength B is called the tesla (T) after the eccentric, but brilliant inventor Nikola Tesla (1856–1943), where

$$1\,T = \frac{1\,N}{A \cdot m}.$$

A smaller unit, called the **gauss** (G) is sometimes used, where

$$1\,G = 10^{-4}\,T$$

The strongest permanent magnets have fields near 2 T; superconducting electromagnets may attain 10 T or more. Earth’s magnetic field on its surface is only about $5 \times 10^{-5}\,T$ or $0.5\,G$.

Problem-Solving Strategy: Direction of the Magnetic Field by the Right-Hand Rule

The direction of the magnetic force $\vec{F}$ is perpendicular to the plane formed by $\vec{v}$ and $\vec{B}$ as determined by the **right-hand rule-1** (or RHR-1), which is illustrated in Figure $6.3.1$.

1. Orient your right hand so that your fingers curl in the plane defined by the velocity and magnetic field vectors.
2. Using your right hand, sweep from the velocity toward the magnetic field with your fingers through the smallest angle possible.
3. The magnetic force is directed where your thumb is pointing.
4. If the charge was negative, reverse the direction found by these steps.

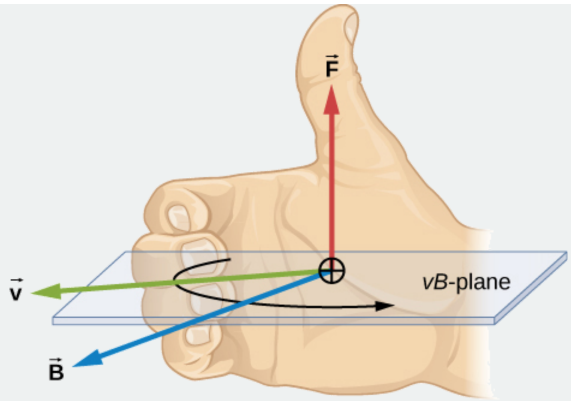


Figure 6.3.1: Magnetic fields exert forces on moving charges. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by b$\vec{v}$ and $\vec{B}$ and follows the right-hand rule-1 (RHR-1) as shown. The magnitude of the force is proportional to q, v, B, and the sine of the angle between $\vec{v}$ and $\vec{B}$.

Note

Visit this website for additional practice with the direction of magnetic fields.

There is no magnetic force on static charges. However, there is a magnetic force on charges moving at an angle to a magnetic field. When charges are stationary, their electric fields do not affect magnets. However, when charges move, they produce magnetic fields that exert forces on other magnets. When there is relative motion, a connection between electric and magnetic forces emerges - each affects the other.

Example 6.3.1: An Alpha-Particle Moving in a Magnetic Field

An alpha-particle $(q = 3.2 \times 10^{-19} C)$ moves through a uniform magnetic field whose magnitude is 1.5 T. The field is directly parallel to the positive **z**-axis of the rectangular coordinate system of Figure 6.3.2. What is the magnetic force on the alpha-particle when it is moving (a) in the positive **x**-direction with a speed of $5.0 \times 10^4 m/s$? (b) in the negative **y**-direction with a speed of $5.0 \times 10^4 m/s$? (c) in the positive **z**-direction with a speed of $5.0 \times 10^4 m/s$? (d) with a velocity $\vec{v} = \left(2.0\hat{i} - 3.0\hat{j} + 1.0\hat{k}\right) \times 10^4 m/s$?

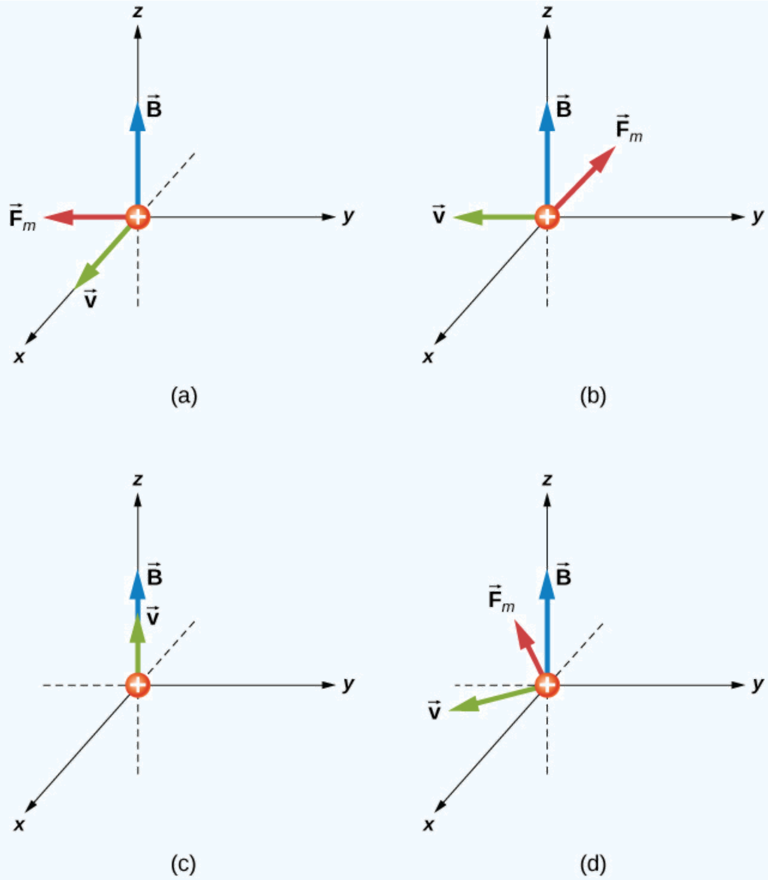


Figure 6.3.2: The magnetic forces on an alpha-particle moving in a uniform magnetic field. The field is the same in each drawing, but the velocity is different.

Strategy

We are given the charge, its velocity, and the magnetic field strength and direction. We can thus use the equation $\vec{F} = q\vec{v} \times \vec{B}$ or $F = qvB sin\,\theta$ to calculate the force. The direction of the force is determined by RHR-1.

Solution

1. First, to determine the direction, start with your fingers pointing in the positive **x**-direction. Sweep your fingers upward in the direction of magnetic field. Your thumb should point in the negative **y**-direction. This should match the mathematical answer. To calculate the force, we use the given charge, velocity, and magnetic field and the definition of the magnetic force in cross-product form to calculate:

$$\vec{F} = q\vec{v} \times \vec{B} = (3.2 \times 10^{-19} C)(5.0 \times 10^4 m/s\,\hat{i}) \times (1.5\,T\,\hat{k}) = -2.4 \times 10^{-14} N\,\hat{j}$$

2. First, to determine the directionality, start with your fingers pointing in the negative **y**-direction. Sweep your fingers upward in the direction of magnetic field as in the previous problem. Your thumb should be open in the negative **x**-direction. This should match the mathematical answer. To calculate the force, we use the given charge, velocity, and magnetic field and the definition of the magnetic force in cross-product form to calculate:

$$\vec{F} = q\vec{v} \times \vec{B} = (3.2 \times 10^{-19} C)(-5.0 \times 10^4 m/s\,\hat{i}) \times (1.5\,T\,\hat{k}) = -2.4 \times 10^{-14} N\,\hat{i}$$

An alternative approach is to use Equation 6.3.2 to find the magnitude of the force. This applies for both parts (a) and (b). Since the velocity is perpendicular to the magnetic field, the angle between them is 90 degrees. Therefore, the magnitude of the force is:

$$F = qvB\sin\,\theta = (3.2 \times 10^{-19} C)(5.0 \times 10^4 m/s)(1.5\,T) sin(90^o)) = 2.4 \times 10^{-14} N.$$

3. Since the velocity and magnetic field are parallel to each other, there is no orientation of your hand that will result in a force direction. Therefore, the force on this moving charge is zero. This is confirmed by the cross product. When you cross two vectors pointing in the same direction, the result is equal to zero.
4. First, to determine the direction, your fingers could point in any orientation; however, you must sweep your fingers upward in the direction of the magnetic field. As you rotate your hand, notice that the thumb can point in any **x**- or **y**-direction possible, but not in the **z**-direction. This should match the mathematical answer. To calculate the force, we use the given charge, velocity, and magnetic field and the definition of the magnetic force in cross-product form to calculate:

$$\vec{F} = q\vec{v} \times \vec{B} = (3.2 \times 10^{-19} C)((2.0\hat{i} - 3.0\hat{j} + 1.0\hat{k}) \times 10^4 m/s) \times (1.5\,T\hat{k})$$

$$(-14.4\hat{i} - 9.6\hat{j}) \times 10^{-15} N.$$

This solution can be rewritten in terms of a magnitude and angle in the **xy**-plane:

$$|\vec{F}| = \sqrt{F_x^2 + F_y^2} = \sqrt{(-14.4)^2 + (-9.6)^2} \times 10^{-15} N = 1.7 \times 10^{-14} N$$

$$\theta = tan^{-1}\left(\frac{F_y}{F_x}\right) = tan^{-1}\left(\frac{-9.6 \times 10^{-15} N}{-14.4 \times 10^{-15} N}\right) = 34^o.$$

The magnitude of the force can also be calculated using Equation 6.3.2. The velocity in this question, however, has three components. The **z**-component of the velocity can be neglected, because it is parallel to the magnetic field and therefore generates no force. The magnitude of the velocity is calculated from the **x**- and **y**-components. The angle between the velocity in the **xy**-plane and the magnetic field in the **z**-plane is 90 degrees. Therefore, the force is calculated to be:

$$|\vec{v}| = \sqrt{(2)^2 + (-3)^2} \times 10^4 \frac{m}{s} = 3.6 \times 10^4 \frac{m}{s}$$

$$F = qvBsin\,\theta = (3.2 \times 10^{-19} C)(3.6 \times 10^4 m/s)(1.5\,T)sin(90^o) = 1.7 \times 10^{-14} N$$

This is the same magnitude of force calculated by unit vectors.

Significance

The cross product in this formula results in a third vector that must be perpendicular to the other two. Other physical quantities, such as angular momentum, also have three vectors that are related by the cross product. Note that typical force values in magnetic force problems are much larger than the gravitational force. Therefore, for an isolated charge, the magnetic force is the dominant force governing the charge's motion.

? Exercise 6.3.1

Repeat the previous problem with the magnetic field in the **x**-direction rather than in the **z**-direction. Check your answers with RHR-1.

Answer a

0 N

Answer b

$2.4 \times 10^{-14}\hat{k}N$

Answer c

$2.4 \times 10^{-14}\hat{j}N$

Answer d

$7.2\hat{j} + 2.2\hat{k}) \times 10^{-15} N$

Representing Magnetic Fields

The representation of magnetic fields by **magnetic field lines** is very useful in visualizing the strength and direction of the magnetic field. As shown in Figure $6.3.3$, each of these lines forms a closed loop, even if not shown by the constraints of the space available for the figure. The field lines emerge from the north pole (N), loop around to the south pole (S), and continue through the bar magnet back to the north pole.

Magnetic field lines have several hard-and-fast rules:

1. The direction of the magnetic field is tangent to the field line at any point in space. A small compass will point in the direction of the field line.
2. The strength of the field is proportional to the closeness of the lines. It is exactly proportional to the number of lines per unit area perpendicular to the lines (called the areal density).
3. Magnetic field lines can never cross, meaning that the field is unique at any point in space.
4. Magnetic field lines are continuous, forming closed loops without a beginning or end. They are directed from the north pole to the south pole.

The last property is related to the fact that the north and south poles cannot be separated. It is a distinct difference from electric field lines, which generally begin on positive charges and end on negative charges or at infinity. If isolated magnetic charges (referred to as **magnetic monopoles**) existed, then magnetic field lines would begin and end on them.

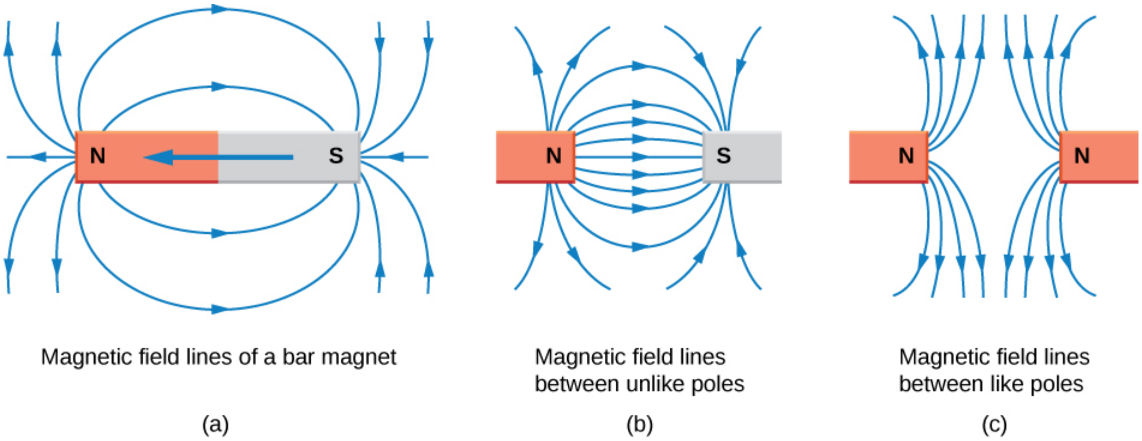


Figure $6.3.3$: Magnetic field lines are defined to have the direction in which a small compass points when placed at a location in the field. The strength of the field is proportional to the closeness (or density) of the lines. If the interior of the magnet could be probed, the field lines would be found to form continuous, closed loops. To fit in a reasonable space, some of these drawings may not show the closing of the loops; however, if enough space were provided, the loops would be closed.

6.4: Force and Torque on a Current Loop

Learning Objectives

By the end of this section, you will be able to:

- Evaluate the net force on a current loop in an external magnetic field
- Evaluate the net torque on a current loop in an external magnetic field
- Define the magnetic dipole moment of a current loop

Motors are the most common application of magnetic force on current-carrying wires. Motors contain loops of wire in a magnetic field. When current is passed through the loops, the magnetic field exerts torque on the loops, which rotates a shaft. Electrical energy is converted into mechanical work in the process. Once the loop's surface area is aligned with the magnetic field, the direction of current is reversed, so there is a continual torque on the loop (Figure 6.4.1). This reversal of the current is done with commutators and brushes. The commutator is set to reverse the current flow at set points to keep continual motion in the motor. A basic commutator has three contact areas to avoid and dead spots where the loop would have zero instantaneous torque at that point. The brushes press against the commutator, creating electrical contact between parts of the commutator during the spinning motion.

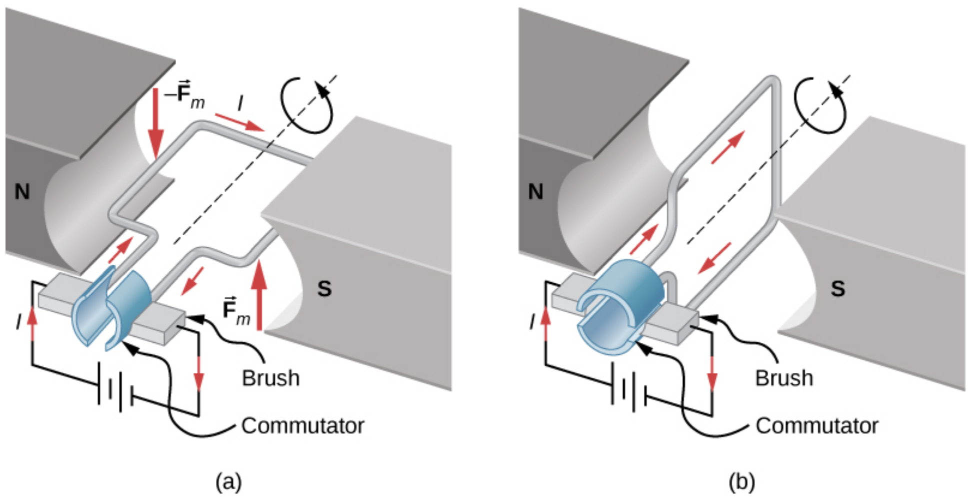


Figure 6.4.1: A simplified version of a dc electric motor. (a) The rectangular wire loop is placed in a magnetic field. The forces on the wires closest to the magnetic poles (N and S) are opposite in direction as determined by the right-hand rule-1. Therefore, the loop has a net torque and rotates to the position shown in (b). (b) The brushes now touch the commutator segments so that no current flows through the loop. No torque acts on the loop, but the loop continues to spin from the initial velocity given to it in part (a). By the time the loop flips over, current flows through the wires again but now in the opposite direction, and the process repeats as in part (a). This causes continual rotation of the loop.

In a uniform magnetic field, a current-carrying loop of wire, such as a loop in a motor, experiences both forces and torques on the loop. Figure 6.4.1 shows a rectangular loop of wire that carries a current **I** and has sides of lengths **a** and **b**. The loop is in a uniform magnetic field: $\vec{B} = B\hat{j}$. The magnetic force on a straight current-carrying wire of length **l** is given by $I\vec{l} \times \vec{B}$. To find the net force on the loop, we have to apply this equation to each of the four sides. The force on side 1 is

$$\vec{F}_1 = IaB\sin(90^o - \theta)\hat{i} = IaB\cos\theta\hat{i}$$

where the direction has been determined with the RHR-1. The current in side 3 flows in the opposite direction to that of side 1, so

$$\vec{F}_3 = -IaB\sin(90^o + \theta)\hat{i} = -IaB\cos\theta\hat{i}$$

The currents in sides 2 and 4 are perpendicular to $\vec{B}$ and the forces on these sides are

$$\vec{F}_2 = IbB\hat{k}$$

$$\vec{F}_4 = -IbB\hat{k}.$$

We can now find the net force on the loop:

$$\sum \vec{F}_{net} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 = 0.$$

Although this result $(\sum F = 0)$ has been obtained for a rectangular loop, it is far more general and holds for current-carrying loops of arbitrary shapes; that is, there is no net force on a current loop in a uniform magnetic field.

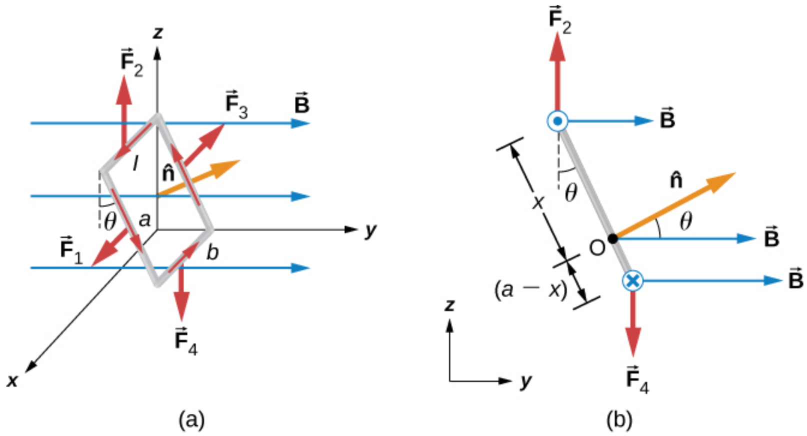


Figure 6.4.2: (a) A rectangular current loop in a uniform magnetic field is subjected to a net torque but not a net force. (b) A side view of the coil.

To find the net torque on the current loop shown in Figure $6.4.2a$, we first consider F_1 and F_3. Since they have the same line of action and are equal and opposite, the sum of their torques about any axis is zero (see Fixed-Axis Rotation). Thus, if there is any torque on the loop, it must be furnished by F_2 and F_4. Let's calculate the torques around the axis that passes through point **O** of Figure $6.4.2b$ (a side view of the coil) and is perpendicular to the plane of the page. The point **O** is a distance **x** from side 2 and a distance $(a-x)$ from side 4 of the loop. The moment arms of F_2 and F_4 are $x\ sin\,\theta$ and $(a-x)\ sin\,\theta$, respectively, so the net torque on the loop is

$$\sum \vec{\tau} = \vec{\tau}_1 + \vec{\tau}_2 + \vec{\tau}_3 + \vec{\tau}_4 = F_2 x\ sin\,\theta \hat{i} - F_4(a-x)\ sin\,(\theta)\hat{i}$$

$$-IbBx\ sin\,\theta \hat{i} - IbB(a-x) sin\,\theta \hat{i}.$$

This simplifies to

$$\vec{\tau} = -IAB\,sin\,\theta \hat{i}$$

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

Notice that this torque is independent of **x**; it is therefore independent of where point **O** is located in the plane of the current loop. Consequently, the loop experiences the same torque from the magnetic field about any axis in the plane of the loop and parallel to the **x**-axis.

A closed-current loop is commonly referred to as a **magnetic dipole** and the term **IA** is known as its **magnetic dipole moment** $\boldsymbol{\mu}$**.** Actually, the magnetic dipole moment is a vector that is defined as

$$\vec{\mu} = IA\hat{n}$$

where $\hat{n}$ is a unit vector directed perpendicular to the plane of the loop (see Figure $6.4.2$). The direction of $\hat{n}$ is obtained with the RHR-2—if you curl the fingers of your right hand in the direction of current flow in the loop, then your thumb points along $\hat{n}$. If the loop contains **N** turns of wire, then its magnetic dipole moment is given by

$$\vec{\mu} = NIA\hat{n}.$$

In terms of the magnetic dipole moment, the torque on a current loop due to a uniform magnetic field can be written simply as

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

This equation holds for a current loop in a two-dimensional plane of arbitrary shape.

Using a calculation analogous to that found in Capacitance for an electric dipole, the potential energy of a magnetic dipole is

$$U = -\vec{\mu} \cdot \vec{B}.$$

Example 6.4.1: Forces and Torques on Current-Carrying Loops

A circular current loop of radius 2.0 cm carries a current of 2.0 mA. (a) What is the magnitude of its magnetic dipole moment? (b) If the dipole is oriented at 30 degrees to a uniform magnetic field of magnitude 0.50 T, what is the magnitude of the torque it experiences and what is its potential energy?

Strategy

The dipole moment is defined by the current times the area of the loop. The area of the loop can be calculated from the area of the circle. The torque on the loop and potential energy are calculated from identifying the magnetic moment, magnetic field, and angle oriented in the field.

Solution

1. The magnetic moment **μ** is calculated by the current times the area of the loop or πr^2.

$$\mu = IA = (2.0 \times 10^{-3} A)(\pi (0.02\, m)^2) = 2.5 \times 10^{-6} A \cdot m^2$$

2. The torque and potential energy are calculated by identifying the magnetic moment, magnetic field, and the angle between these two vectors. The calculations of these quantities are:

$$\tau = \vec{\mu} \times \vec{B} = \mu B\, sin\, \theta = (2.5 \times 10^{-6} A \cdot m^2)(0.50 T) sin(30^o) = 6.3 \times 10^{-7} N \cdot m$$

$$U = -\vec{\mu} \cdot \vec{B} = -\mu B cos\theta = -(2.5 \times 10^{-6} A \cdot m^2)(0.50 T) cos(30^o) = -1.1 \times 10^{-6} J.$$

Significance

The concept of magnetic moment at the atomic level is discussed in the next chapter. The concept of aligning the magnetic moment with the magnetic field is the functionality of devices like magnetic motors, whereby switching the external magnetic field results in a constant spinning of the loop as it tries to align with the field to minimize its potential energy.

Exercise 6.4.1

In what orientation would a magnetic dipole have to be to produce (a) a maximum torque in a magnetic field? (b) A maximum energy of the dipole?

Solution

a. aligned or anti-aligned; b. perpendicular

6.5: The Hall Effect

Learning Objectives

By the end of this section, you will be able to:

- Explain a scenario where the magnetic and electric fields are crossed and their forces balance each other as a charged particle moves through a velocity selector
- Compare how charge carriers move in a conductive material and explain how this relates to the Hall effect

In 1879, E.H. Hall devised an experiment that can be used to identify the sign of the predominant charge carriers in a conducting material. From a historical perspective, this experiment was the first to demonstrate that the charge carriers in most metals are negative.

Visit this website to find more information about the Hall effect.

We investigate the **Hall effect** by studying the motion of the free electrons along a metallic strip of width **l** in a constant magnetic field (Figure 6.5.1). The electrons are moving from left to right, so the magnetic force they experience pushes them to the bottom edge of the strip. This leaves an excess of positive charge at the top edge of the strip, resulting in an electric field **E** directed from top to bottom. The charge concentration at both edges builds up until the electric force on the electrons in one direction is balanced by the magnetic force on them in the opposite direction. Equilibrium is reached when:

$$eE = ev_d B \tag{6.5.1}$$

where **e** is the magnitude of the electron charge, v_d is the drift speed of the electrons, and **E** is the magnitude of the electric field created by the separated charge. Solving this for the drift speed results in

$$v_d = \frac{E}{B}. \tag{6.5.2}$$

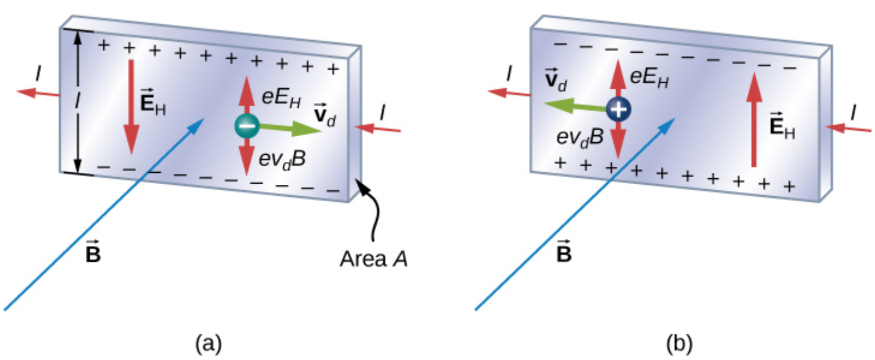


Figure 6.5.1: In the Hall effect, a potential difference between the top and bottom edges of the metal strip is produced when moving charge carriers are deflected by the magnetic field. (a) Hall effect for negative charge carriers; (b) Hall effect for positive charge carriers.

A scenario where the electric and magnetic fields are perpendicular to one another is called a crossed-field situation. If these fields produce equal and opposite forces on a charged particle with the velocity that equates the forces, these particles are able to pass through an apparatus, called a **velocity selector**, undeflected. This velocity is represented in Equation 6.5.3. Any other velocity of a charged particle sent into the same fields would be deflected by the magnetic force or electric force.

Going back to the Hall effect, if the current in the strip is **I**, then from Current and Resistance, we know that

$$I = nev_d A \tag{6.5.3}$$

where **n** is the number of charge carriers per volume and **A** is the cross-sectional area of the strip. Combining the equations for v_d and **I** results in

$$I = ne\left(\frac{E}{B}\right)A. \tag{6.5.4}$$

The field **E** is related to the potential difference **V** between the edges of the strip by

$$E = \frac{V}{l}. \tag{6.5.5}$$

The quantity V is called the **Hall potential** and can be measured with a voltmeter. Finally, combining the equations for **I** and **E** gives us

$$V = \frac{IBl}{neA} \tag{6.5.6}$$

where the upper edge of the strip in Figure $6.5.1$ is positive with respect to the lower edge.

We can also combine Equation 6.5.1 and Equation 6.5.5 to get an expression for the Hall voltage in terms of the magnetic field:

$$V = Blv_d.$$

What if the charge carriers are positive, as in Figure $6.5.1$? For the same current **I**, the magnitude of **V** is still given by Equation 6.5.6. However, the upper edge is now negative with respect to the lower edge. Therefore, by simply measuring the sign of **V**, we can determine the sign of the majority charge carriers in a metal.

Hall potential measurements show that electrons are the dominant charge carriers in most metals. However, Hall potentials indicate that for a few metals, such as tungsten, beryllium, and many semiconductors, the majority of charge carriers are positive. It turns out that conduction by positive charge is caused by the migration of missing electron sites (called holes) on ions. Conduction by holes is studied later in Condensed Matter Physics.

The Hall effect can be used to measure magnetic fields. If a material with a known density of charge carriers **n** is placed in a magnetic field and **V** is measured, then the field can be determined from Equation ???. In research laboratories where the fields of electromagnets used for precise measurements have to be extremely steady, a “Hall probe” is commonly used as part of an electronic circuit that regulates the field.

✔ Example 6.5.1: Velocity Selector

An electron beam enters a crossed-field velocity selector with magnetic and electric fields of 2.0 mT and $6.0 \times 10^3\, N/C$, respectively. (a) What must the velocity of the electron beam be to traverse the crossed fields undeflected? If the electric field is turned off, (b) what is the acceleration of the electron beam and (c) what is the radius of the circular motion that results?

Strategy

The electron beam is not deflected by either of the magnetic or electric fields if these forces are balanced. Based on these balanced forces, we calculate the velocity of the beam. Without the electric field, only the magnetic force is used in Newton’s second law to find the acceleration. Lastly, the radius of the path is based on the resulting circular motion from the magnetic force.

Solution

1. The velocity of the unperturbed beam of electrons with crossed fields is calculated by Equation 6.5.2:

$$v_d = \frac{E}{B} = \frac{6 \times 10^3 N/C}{2 \times 10^{-3} T} = 3 \times 10^6 m/s.$$

2. The acceleration is calculated from the net force from the magnetic field, equal to mass times acceleration. The magnitude of the acceleration is:

$$ma = qvB$$

$$a = \frac{qvB}{m} = \frac{(1.6 \times 10^{-19} C)(3 \times 10^6 m/s)(2 \times 10^{-3} T)}{0.1 \times 10^{-31} kg} = 1.1 \times 10^{15} m/s^2.$$

3. The radius of the path comes from a balance of the circular and magnetic forces, or Equation 6.5.2:

$$r = \frac{mv}{qB} = \frac{(9.1 \times 10^{-31} kg)(3 \times 10^6 m/s)}{(1.6 \times 10^{-19} C)(2 \times 10^{-3} T)} = 8.5 \times 10^{-3} m.$$

Significance

If electrons in the beam had velocities above or below the answer in part (a), those electrons would have a stronger net force exerted by either the magnetic or electric field. Therefore, only those electrons at this specific velocity would make it through.

✓ The Hall Potential in a Silver Ribbon

Figure $6.5.2$ shows a silver ribbon whose cross section is 1.0 cm by 0.20 cm. The ribbon carries a current of 100 A from left to right, and it lies in a uniform magnetic field of magnitude 1.5 T. Using a density value of $n = 5.9 \times 10^{28}$ electrons per cubic meter for silver, find the Hall potential between the edges of the ribbon.

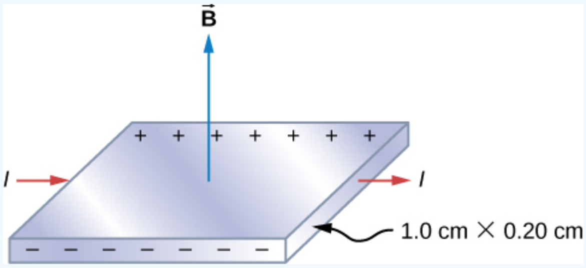


Figure $6.5.2$: Finding the Hall potential in a silver ribbon in a magnetic field is shown.

Strategy

Since the majority of charge carriers are electrons, the polarity of the Hall voltage is that indicated in the figure. The value of the Hall voltage is calculated using Equation 6.5.6.

Solution

When calculating the Hall voltage, we need to know the current through the material, the magnetic field, the length, the number of charge carriers, and the area. Since all of these are given, the Hall voltage is calculated as:

$$\begin{aligned} v &= \frac{IBl}{neA} \\ &= \frac{(100\,A)(1.5\,T)(1.0 \times 10^{-2} m)}{(5.9 \times 10^{28}/m^3)(1.6 \times 10^{-19} C)(2.0 \times 10^{-5} m^2)} \\ &= 7.9 \times 10^{-6} V. \end{aligned}$$

Significance

As in this example, the Hall potential is generally very small, and careful experimentation with sensitive equipment is required for its measurement.

? Exercise 6.5.1

A Hall probe consists of a copper strip, $n = 8.5 \times 10^{28}$ electrons per cubic meter, which is 2.0 cm wide and 0.10 cm thick. What is the magnetic field when **I** = 50 A and the Hall potential is

a. $4.0\,\mu V$ and
b. $6.0\,\mu V$?

Answer a

1.1 T

Answer b

1.6 T

CHAPTER OVERVIEW

7: Electromagnetic Waves

In this chapter, we explain Maxwell's theory and show how it leads to his prediction of electromagnetic waves. We use his theory to examine what electromagnetic waves are, how they are produced, and how they transport energy and momentum. We conclude by summarizing some of the many practical applications of electromagnetic waves.

7.1: Prelude to Electromagnetic Waves

Our view of objects in the sky at night, the warm radiance of sunshine, the sting of sunburn, our cell phone conversations, and the X-rays revealing a broken bone—all are brought to us by electromagnetic waves. It would be hard to overstate the practical importance of electromagnetic waves, through their role in vision, through countless technological applications, and through their ability to transport the energy from the Sun through space to sustain life and almost all of its activities on Earth.

Figure 7.1.16: The pressure from sunlight predicted by Maxwell's equations helped produce the tail of Comet McNaught. (credit: modification of work by Sebastian Deiries—ESO)

Theory predicted the general phenomenon of electromagnetic waves before anyone realized that light is a form of an electromagnetic wave. In the mid-nineteenth century, James Clerk Maxwell formulated a single theory combining all the electric and magnetic effects known at that time. Maxwell's equations, summarizing this theory, predicted the existence of electromagnetic waves that travel at the speed of light. His theory also predicted how these waves behave, and how they carry both energy and momentum. The tails of comets, such as Comet McNaught in Figure 16.1, provide a spectacular example. Energy carried by light from the Sun warms the comet to release dust and gas. The momentum carried by the light exerts a weak force that shapes the dust into a tail of the kind seen here. The flux of particles emitted by the Sun, called the solar wind, typically produces an additional, second tail, as described in detail in this chapter.

In this chapter, we explain Maxwell's theory and show how it leads to his prediction of electromagnetic waves. We use his theory to examine what electromagnetic waves are, how they are produced, and how they transport energy and momentum. We conclude by summarizing some of the many practical applications of electromagnetic waves.

7.2: Maxwell's Equations and Electromagnetic Waves

Learning Objectives

By the end of this section, you will be able to:

- Explain Maxwell's correction of Ampère's law by including the displacement current
- State and apply Maxwell's equations in integral form
- Describe how the symmetry between changing electric and changing magnetic fields explains Maxwell's prediction of electromagnetic waves
- Describe how Hertz confirmed Maxwell's prediction of electromagnetic waves

James Clerk **Maxwell** (1831–1879) was one of the major contributors to physics in the nineteenth century (Figure $7.2.1$). Although he died young, he made major contributions to the development of the kinetic theory of gases, to the understanding of color vision, and to the nature of Saturn's rings. He is probably best known for having combined existing knowledge of the laws of electricity and of magnetism with insights of his own into a complete overarching electromagnetic theory, represented by **Maxwell's equations**.

Figure $7.2.1$: James Clerk Maxwell, a nineteenth-century physicist, developed a theory that explained the relationship between electricity and magnetism, and correctly predicted that visible light consists of electromagnetic waves.

Maxwell's Correction to the Laws of Electricity and Magnetism

The four basic laws of electricity and magnetism had been discovered experimentally through the work of physicists such as Oersted, Coulomb, Gauss, and Faraday. Maxwell discovered logical inconsistencies in these earlier results and identified the incompleteness of Ampère's law as their cause.

Recall that according to Ampère's law, the integral of the magnetic field around a closed loop **C** is proportional to the current **I** passing through any surface whose boundary is loop **C** itself:

$$\oint \vec{B} \cdot d\vec{s} = \mu_0 I. \tag{7.2.1}$$

There are infinitely many surfaces that can be attached to any loop, and Ampère's law stated in Equation 7.2.1 is independent of the choice of surface.

Consider the set-up in Figure $7.2.2$. A source of emf is abruptly connected across a parallel-plate capacitor so that a time-dependent current **I** develops in the wire. Suppose we apply Ampère's law to loop **C** shown at a time before the capacitor is fully charged, so that $I \neq 0$. Surface S_1 gives a nonzero value for the enclosed current **I**, whereas surface S_2 gives zero for the enclosed current because no current passes through it:

$$\underbrace{\oint_C \vec{B}\cdot d\vec{s} = \mu_0 I}_{\text{if surface } S_1 \text{is used}}$$

$$\underbrace{=0}_{\text{if surface } S_2 \text{is used}}$$

Clearly, Ampère’s law in its usual form does not work here. This may not be surprising, because Ampère’s law as applied in earlier chapters required a steady current, whereas the current in this experiment is changing with time and is not steady at all.

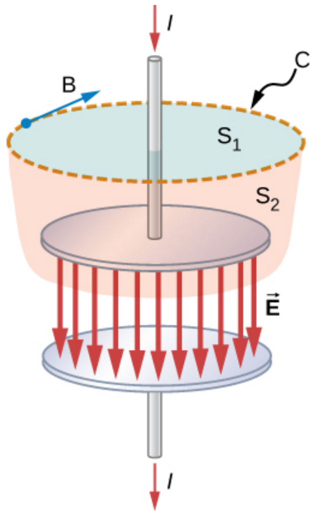


Figure 7.2.2: The currents through surface S_1 and surface S_2 are unequal, despite having the same boundary loop **C**.

How can Ampère’s law be modified so that it works in all situations? Maxwell suggested including an additional contribution, called the displacement current I_d, to the real current **I**,

$$\boxed{\oint_S \vec{B}\cdot d\vec{s} = \mu_0(I+I_d)} \tag{7.2.2}$$

where the displacement current is defined to be

$$\boxed{I_d = \epsilon_0 \frac{d\Phi_E}{dt}.} \tag{7.2.3}$$

Here ϵ_0 is the **permittivity of free space** and Φ_E is the electric flux, defined as

$$\Phi_E = \iint_{Surface\ S} \vec{E}\cdot d\vec{A}.$$

The **displacement current** is analogous to a real current in Ampère’s law, entering into Ampère’s law in the same way. It is produced, however, by a changing electric field. It accounts for a changing electric field producing a magnetic field, just as a real current does, but the displacement current can produce a magnetic field even where no real current is present. When this extra term is included, the modified Ampère’s law equation becomes

$$\oint_C \vec{B}\cdot d\vec{s} = \mu_0 I + \epsilon_0\mu_0 \frac{d\Phi_E}{dt}$$

and is independent of the surface **S** through which the current **I** is measured.

We can now examine this modified version of Ampère’s law to confirm that it holds independent of whether the surface S_1 or the surface S_2 in Figure 7.2.2 is chosen. The electric field $\vec{E}$ corresponding to the flux Φ_E in Equation 7.2.3 is between the capacitor plates. Therefore, the $\vec{E}$ field and the displacement current through the surface S_1 are both zero, and Equation 7.2.2 takes the form

$$\oint_C \vec{B}\cdot d\vec{s} = \mu_0 I. \tag{7.2.4}$$

We must now show that for surface S_2, through which no actual current flows, the displacement current leads to the same value $\mu_0 I$ for the right side of the Ampère's law equation. For surface S_2 the equation becomes

$$\oint_C \vec{B}\cdot d\vec{s} = \mu_0 \frac{d}{dt}\left[\epsilon_0 \iint_{Surface\ S_2} \vec{E}\cdot d\vec{A}\right].$$

Gauss's law for electric charge requires a closed surface and cannot ordinarily be applied to a surface like S_1 alone or S_2 alone. But the two surfaces S_1 and S_2 form a closed surface in Figure 7.2.2 and can be used in Gauss's law. Because the electric field is zero on S_1, the flux contribution through S_1 is zero. This gives us

$$\oint_{Surface\ S_1+S_2} \vec{E}\cdot d\vec{A} = \iint_{Surface\ S_1} \vec{E}\cdot d\vec{A} + \iint_{Surface\ S_2} \vec{E}\cdot d\vec{A} \tag{7.2.5}$$

$$= 0 + \iint_{Surface\ S_2} \vec{E}\cdot d\vec{A} \tag{7.2.6}$$

$$= \iint_{Surface\ S_2} \vec{E}\cdot d\vec{A}. \tag{7.2.7}$$

Therefore, we can replace the integral over S_2 in Equation 7.2.4 with the closed Gaussian surface $S_1 + S_2$ and apply Gauss's law to obtain

$$\oint_{S_1} \vec{B}\cdot d\vec{s} = \mu_0 \frac{dQ_{in}}{dt} = \mu_0 I.$$

Thus, the modified Ampère's law equation is the same using surface S_2, where the right-hand side results from the displacement current, as it is for the surface S_1, where the contribution comes from the actual flow of electric charge.

✔ Displacement current in a charging capacitor

A parallel-plate capacitor with capacitance **C** whose plates have area **A** and separation distance **d** is connected to a resistor **R** and a battery of voltage **V**. The current starts to flow at $t = 0$.

a. Find the displacement current between the capacitor plates at time **t**.
b. From the properties of the capacitor, find the corresponding real current $I = \frac{dQ}{dt}$, and compare the answer to the expected current in the wires of the corresponding **RC** circuit.

Strategy

We can use the equations from the analysis of an **RC** circuit (Alternating-Current Circuits) plus Maxwell's version of Ampère's law.

Solution

a. The voltage between the plates at time **t** is given by

$$V_C = \frac{1}{C}Q(t) = V_0\left(1 - e^{-t/RC}\right).$$

Let the **z**-axis point from the positive plate to the negative plate. Then the **z**-component of the electric field between the plates as a function of time **t** is

$$E_z(t) = \frac{V_0}{d}\left(1 - e^{-t/RC}\right).$$

Therefore, the z-component of the displacement current I_d between the plates is

$$I_d(t) = \epsilon_0 A \frac{\partial E_z(t)}{\partial t} = \epsilon_0 A \frac{V_0}{d} \times \frac{1}{RC} e^{-t/RC} = \frac{V_0}{R} e^{-t/RC},$$

where we have used $C = \epsilon_0 \frac{A}{d}$ for the capacitance.

b. From the expression for V_C the charge on the capacitor is

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

The current into the capacitor after the circuit is closed, is therefore

$$I = \frac{dQ}{dt} = \frac{V_0}{R} e^{-t/RC}.$$

This current is the same as I_d found in (a).

Maxwell's Equations

With the correction for the displacement current, Maxwell's equations take the form

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q_{in}}{\epsilon_0} \text{ (Gauss's law)} \tag{7.2.8}$$

$$\oint \vec{B} \cdot d\vec{A} = 0 \text{ (Gauss's law for magnetism)} \tag{7.2.9}$$

$$\oint \vec{E} \cdot d\vec{s} = -\frac{d\Phi_m}{dt} \text{ (Faraday's law)} \tag{7.2.10}$$

$$\oint \vec{B} \cdot d\vec{s} = \mu_0 I + \epsilon_0 I + \epsilon_0 \mu_0 \frac{d\Phi_E}{dt} \text{(Ampere-Maxwell law)}. \tag{7.2.11}$$

Once the fields have been calculated using these four equations, the **Lorentz force equation**

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$$

gives the force that the fields exert on a particle with charge **q** moving with velocity $\vec{v}$. The Lorentz force equation combines the force of the electric field and of the magnetic field on the moving charge. The magnetic and electric forces have been examined in earlier modules. These four Maxwell's equations are, respectively:

Maxwell's Equations

1. **Gauss's law**

The electric flux through any closed surface is equal to the electric charge Q_{in} enclosed by the surface. Gauss's law (Equation 7.2.8) describes the relation between an electric charge and the electric field it produces. This is often pictured in terms of electric field lines originating from positive charges and terminating on negative charges, and indicating the direction of the electric field at each point in space.

2. **Gauss's law for magnetism**

The magnetic field flux through any closed surface is zero (Equation 7.2.9). This is equivalent to the statement that magnetic field lines are continuous, having no beginning or end. Any magnetic field line entering the region enclosed by the surface must also leave it. No magnetic monopoles, where magnetic field lines would terminate, are known to exist (see section on Magnetic Fields and Lines).

3. **Faraday's law**

A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. Equation 7.2.10 is Faraday's law of induction and includes Lenz's law. The electric field from a changing magnetic field has field lines that form closed loops, without any beginning or end.

4. **Ampère-Maxwell law**

Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell's equations, Equation 7.2.11, encompasses Ampère's law and adds another source of magnetic fields, namely changing electric fields.

Maxwell's equations and the Lorentz force law together encompass all the laws of electricity and magnetism. The symmetry that Maxwell introduced into his mathematical framework may not be immediately apparent. Faraday's law describes how changing magnetic fields produce electric fields. The displacement current introduced by Maxwell results instead from a changing electric field and accounts for a changing electric field producing a magnetic field. The equations for the effects of both changing electric fields and changing magnetic fields differ in form only where the absence of magnetic monopoles leads to missing terms. This symmetry between the effects of changing magnetic and electric fields is essential in explaining the nature of electromagnetic waves.

Later application of Einstein's theory of relativity to Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate but are different manifestations of the same thing—the electromagnetic force. The electromagnetic force and weak nuclear force are similarly unified as the electroweak force. This unification of forces has been one motivation for attempts to unify all of the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces (see Particle Physics and Cosmology).

The Mechanism of Electromagnetic Wave Propagation

To see how the symmetry introduced by Maxwell accounts for the existence of combined electric and magnetic waves that propagate through space, imagine a time-varying magnetic field $\vec{B}_0(t)$ produced by the high-frequency alternating current seen in Figure $7.2.3$. We represent $\vec{B}_0(t)$ in the diagram by one of its field lines. From Faraday's law, the changing magnetic field through a surface induces a time-varying electric field $\vec{E}_0(t)$ at the boundary of that surface. The displacement current source for the electric field, like the Faraday's law source for the magnetic field, produces only closed loops of field lines, because of the mathematical symmetry involved in the equations for the induced electric and induced magnetic fields. A field line representation of $\vec{E}_0(t)$ is shown. In turn, the changing electric field $\vec{E}_0(t)$ creates a magnetic field $\vec{B}_1(t)$ according to the modified Ampère's law. This changing field induces $\vec{E}_1(t)$ which induces $\vec{B}_2(t)$ and so on. We then have a self-continuing process that leads to the creation of time-varying electric and magnetic fields in regions farther and farther away from **O**. This process may be visualized as the propagation of an electromagnetic wave through space.

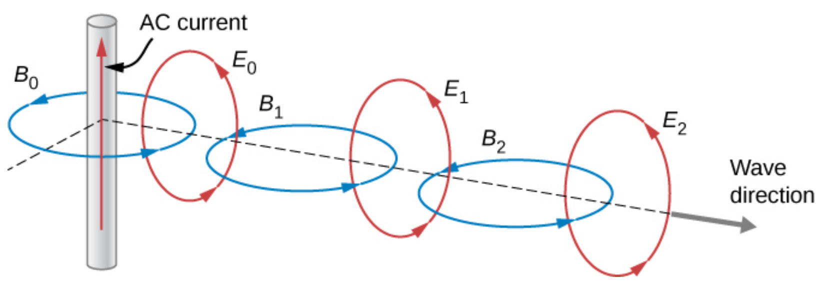


Figure $7.2.3$: How changing $\vec{E}$ and $\vec{B}$ fields propagate through space.

In the next section, we show in more precise mathematical terms how Maxwell's equations lead to the prediction of electromagnetic waves that can travel through space without a material medium, implying a speed of electromagnetic waves equal to the speed of light.

Prior to Maxwell's work, experiments had already indicated that light was a wave phenomenon, although the nature of the waves was yet unknown. In 1801, Thomas Young (1773–1829) showed that when a light beam was separated by two narrow slits and then recombined, a pattern made up of bright and dark fringes was formed on a screen. Young explained this behavior by assuming that light was composed of waves that added constructively at some points and destructively at others (see Interference). Subsequently, Jean Foucault (1819–1868), with measurements of the speed of light in various media, and Augustin Fresnel (1788–1827), with detailed experiments involving interference and diffraction of light, provided further conclusive evidence that light was a wave. So, light was known to be a wave, and Maxwell had predicted the existence of electromagnetic waves that traveled at the speed of light. The conclusion seemed inescapable: Light must be a form of electromagnetic radiation. But Maxwell's theory showed that other wavelengths and frequencies than those of light were possible for electromagnetic waves. He showed that electromagnetic radiation with the same fundamental properties as visible light should exist at any frequency. It remained for others to test, and confirm, this prediction.

? Exercise 7.2.1

When the emf across a capacitor is turned on and the capacitor is allowed to charge, when does the magnetic field induced by the displacement current have the greatest magnitude?

Solution

It is greatest immediately after the current is switched on. The displacement current and the magnetic field from it are proportional to the rate of change of electric field between the plates, which is greatest when the plates first begin to charge.

Hertz's Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves but also verified that they travel at the speed of light.

Hertz used an alternating-current **RLC** (resistor-inductor-capacitor) circuit that resonates at a known frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and connected it to a loop of wire, as shown in Figure 7.2.4. High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and helped generate electromagnetic waves.

Across the laboratory, Hertz placed another loop attached to another **RLC** circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could thus be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.

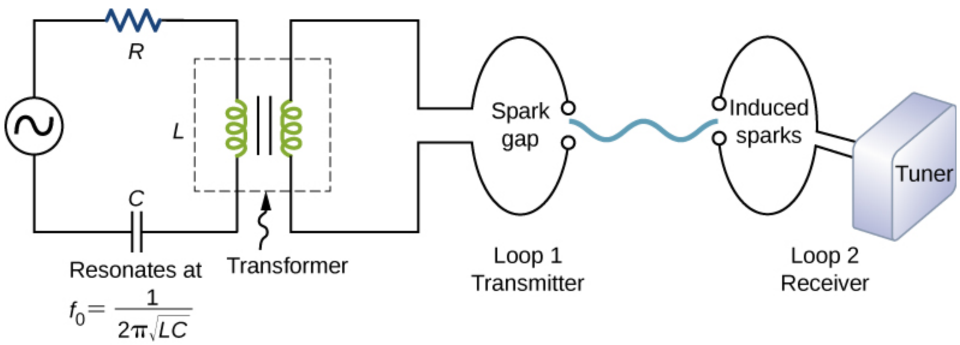


Figure 7.2.4: The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, confirming their wave character. He was able to determine the wavelengths from the interference patterns, and knowing their frequencies, he could calculate the propagation speed using the equation $v = f\lambda$, where **v** is the speed of a wave, **f** is its frequency, and λ is its wavelength. Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz $(1\ Hz = 1\ cycle/second)$, is named in his honor.

? Exercise 7.2.2

Could a purely electric field propagate as a wave through a vacuum without a magnetic field? Justify your answer.

Solution

No. The changing electric field according to the modified version of Ampère's law would necessarily induce a changing magnetic field.

7.3: Plane Electromagnetic Waves

> **LEARNING OBJECTIVES**
>
> By the end of this section, you will be able to:
>
> - Describe how Maxwell's equations predict the relative directions of the electric fields and magnetic fields, and the direction of propagation of plane electromagnetic waves
> - Explain how Maxwell's equations predict that the speed of propagation of electromagnetic waves in free space is exactly the speed of light
> - Calculate the relative magnitude of the electric and magnetic fields in an electromagnetic plane wave
> - Describe how electromagnetic waves are produced and detected

Mechanical waves travel through a medium such as a string, water, or air. Perhaps the most significant prediction of Maxwell's equations is the existence of combined electric and magnetic (or electromagnetic) fields that propagate through space as electromagnetic waves. Because Maxwell's equations hold in free space, the predicted electromagnetic waves, unlike mechanical waves, do not require a medium for their propagation.

A general treatment of the physics of electromagnetic waves is beyond the scope of this textbook. We can, however, investigate the special case of an electromagnetic wave that propagates through free space along the **x**-axis of a given coordinate system.

Electromagnetic Waves in One Direction

An electromagnetic wave consists of an electric field, defined as usual in terms of the force per charge on a stationary charge, and a magnetic field, defined in terms of the force per charge on a moving charge. The electromagnetic field is assumed to be a function of only the **x**-coordinate and time. The **y**-component of the electric field is then written as $E_y(x,t)$, the **z**-component of the magnetic field as $B_z(x,t)$, etc. Because we are assuming free space, there are no free charges or currents, so we can set $Q_{in}=0$ and $I=0$ in Maxwell's equations.

The transverse nature of electromagnetic waves

We examine first what Gauss's law for electric fields implies about the relative directions of the electric field and the propagation direction in an electromagnetic wave. Assume the Gaussian surface to be the surface of a rectangular box whose cross-section is a square of side **l** and whose third side has length Δx, as shown in Figure $7.3.1$. Because the electric field is a function only of **x** and **t**, the **y**-component of the electric field is the same on both the top (labeled Side 2) and bottom (labeled Side 1) of the box, so that these two contributions to the flux cancel. The corresponding argument also holds for the net flux from the **z**-component of the electric field through Sides 3 and 4. Any net flux through the surface therefore comes entirely from the **x**-component of the electric field. Because the electric field has no **y**- or **z**-dependence, $E_x(x,t)$ is constant over the face of the box with area **A** and has a possibly different value $E_x(x+\Delta x,t)$ that is constant over the opposite face of the box.

Applying Gauss's law gives

$$\text{Net flux} = -E_x(x,t)A + E_x(x+\Delta x,t)A = \frac{Q_{in}}{\epsilon_0} \tag{7.3.1}$$

where $A=l\times l$ is the area of the front and back faces of the rectangular surface. But the charge enclosed is $Q_{in}=0$, so this component's net flux is also zero, and Equation 7.3.1 implies $E_x(x,t)=E_x(x+\Delta x,t)$ for any Δx. Therefore, if there is an **x**-component of the electric field, it cannot vary with **x**. A uniform field of that kind would merely be superposed artificially on the traveling wave, for example, by having a pair of parallel-charged plates. Such a component $E_x(x,t)$ would not be part of an electromagnetic wave propagating along the **x**-axis; so $E_x(x,t)=0$ for this wave. Therefore, the only nonzero components of the electric field are $E_y(x,t)$ and $E_z(x,t)$ perpendicular to the direction of propagation of the wave.

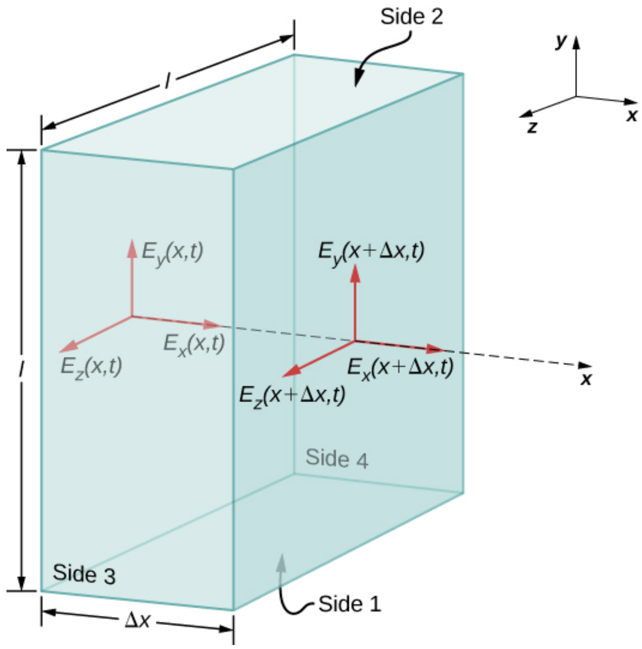


Figure 7.3.1: The surface of a rectangular box of dimensions $l \times l \times \Delta x$ is our Gaussian surface. The electric field shown is from an electromagnetic wave propagating along the x-axis.

A similar argument holds by substituting **E** for **B** and using Gauss's law for magnetism instead of Gauss's law for electric fields. This shows that the **B** field is also perpendicular to the direction of propagation of the wave. The electromagnetic wave is therefore a transverse wave, with its oscillating electric and magnetic fields perpendicular to its direction of propagation.

The speed of propagation of electromagnetic waves

We can next apply Maxwell's equations to the description given in connection with Figure 16.2.3 in the previous section to obtain an equation for the **E** field from the changing **B** field, and for the **B** field from a changing **E** field. We then combine the two equations to show how the changing **E** and **B** fields propagate through space at a speed precisely equal to the speed of light.

First, we apply Faraday's law over Side 3 of the Gaussian surface, using the path shown in Figure 7.3.2. Because $E_x(x,t) = 0$, we have

$$\oint \vec{E} \cdot d\vec{s} = -E_y(x,t)l + E_y(x+\Delta x,t)l.$$

Assuming Δx is small and approximating $E_y(x+\Delta x,t)$ by

$$E_y(x+\Delta x,t) = E_y(x,t) + \frac{\partial E_y(x,t)}{\partial x}\Delta x,$$

we obtain

$$\oint \vec{E} \cdot d\vec{s} = \frac{\partial E_y(x,t)}{\partial x}(l\Delta x).$$

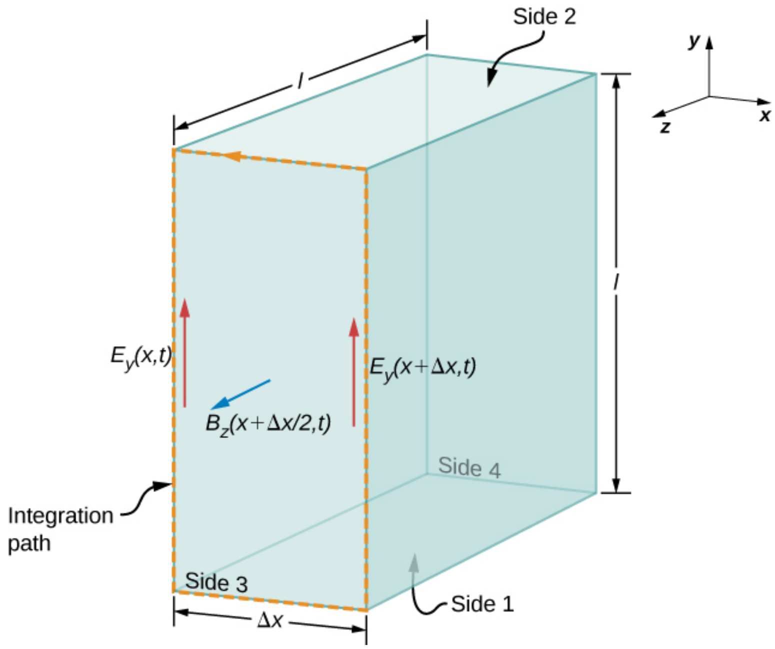


Figure 7.3.2: We apply Faraday's law to the front of the rectangle by evaluating $\oint \vec{E} \cdot d\vec{s}$ along the rectangular edge of Side 3 in the direction indicated, taking the B field crossing the face to be approximately its value in the middle of the area traversed.

Because Δx is small, the magnetic flux through the face can be approximated by its value in the center of the area traversed, namely $B_z\left(x+\frac{\Delta x}{2}, t\right)$. The flux of the **B** field through Face 3 is then the **B** field times the area,

$$\oint_S \vec{B} \cdot \vec{n} dA = B_z\left(x+\frac{\Delta x}{2}, t\right)(l\Delta x). \tag{7.3.2}$$

From Faraday's law,

$$\oint \vec{E} \cdot d\vec{s} = -\frac{d}{dt}\int_S \vec{B} \cdot \vec{n} dA. \tag{7.3.3}$$

Therefore, from Equations 7.3.1 and 7.3.2,

$$\frac{\partial E_y(x,t)}{\partial x}(l\Delta x) = -\frac{\partial}{\partial t}\left[B_z\left(x+\frac{\Delta x}{2}, t\right)\right](l\Delta x).$$

Canceling $l\Delta x$ and taking the limit as $\Delta x = 0$, we are left with

$$\frac{\partial E_y(x,t)}{\partial x} = -\frac{\partial B_z(x,t)}{\partial t}. \tag{7.3.4}$$

We could have applied Faraday's law instead to the top surface (numbered 2) in Figure $7.3.2$, to obtain the resulting equation

$$\frac{\partial B_z(x,t)}{\partial t} = -\frac{\partial E_y(x,t)}{\partial x}. \tag{7.3.5}$$

This is the equation describing the spatially dependent **E** field produced by the time-dependent **B** field.

Next we apply the Ampère-Maxwell law (with $I = 0$) over the same two faces (Surface 3 and then Surface 2) of the rectangular box of Figure $7.3.2$. Applying Equation 16.2.16,

$$\oint \vec{B} \cdot d\vec{s} = \mu_0 \epsilon_0 (d/dt) \int_S \vec{E} \cdot n\, da$$

to Surface 3, and then to Surface 2, yields the two equations

$$\frac{\partial E_y(x,t)}{\partial x} = -\epsilon_0 \mu_0 \frac{\partial E_z(x,t)}{\partial t}, \tag{7.3.6}$$

and

$$\frac{\partial B_z(x,t)}{\partial x} = -\epsilon_0 \mu_0 \frac{\partial E_y(x,t)}{\partial t}. \tag{7.3.7}$$

These equations describe the spatially dependent **B** field produced by the time-dependent **E** field.

We next combine the equations showing the changing **B** field producing an **E** field with the equation showing the changing **E** field producing a **B** field. Taking the derivative of Equation 7.3.4 with respect to **x** and using Equation 7.3.13 gives

$$\frac{\partial^2 E_y}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial E_y}{\partial x}\right) = -\frac{\partial}{\partial x}\left(\frac{\partial B_z}{\partial t}\right) = -\frac{\partial}{\partial t}\left(\frac{\partial B_z}{\partial x}\right) = \frac{\partial}{\partial t}\left(\epsilon_0 \mu_0 \frac{\partial E_y}{\partial t}\right)$$

or

$$\frac{\partial^2 E_y}{\partial x^2} = \epsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2}$$

This is the form taken by the general wave equation for our plane wave. Because the equations describe a wave traveling at some as-yet-unspecified speed **c**, we can assume the field components are each functions of **x – ct** for the wave traveling in the +**x**-direction, that is,

$$E_y(x,t) = f(\xi)\, where\, \xi = x - ct. \tag{7.3.8}$$

It is left as a mathematical exercise to show, using the chain rule for differentiation, that Equations 7.3.5 and 7.3.6 imply

$$1 = \epsilon_0 \mu_0 c^2.$$

The speed of the electromagnetic wave in free space is therefore given in terms of the permeability and the permittivity of free space by

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}. \tag{7.3.9}$$

We could just as easily have assumed an electromagnetic wave with field components $E_z(x,t)$ and $B_y(x,t)$. The same type of analysis with Equation 7.3.12 and 7.3.11 would also show that the speed of an electromagnetic wave is $c = 1/\sqrt{\epsilon_0 \mu_0}$.

The physics of traveling electromagnetic fields was worked out by Maxwell in 1873. He showed in a more general way than our derivation that electromagnetic waves always travel in free space with a speed given by Equation 7.3.6. If we evaluate the speed $c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$, we find that

$$c = \frac{1}{\sqrt{\left(8.85 \times 10^{-12} \frac{C^2}{N \cdot m^2}\right)\left(4\pi \times 10^{-7} \frac{T \cdot m}{A}\right)}} = 3.00 \times 10^8 m/s,$$

which is the speed of light. Imagine the excitement that Maxwell must have felt when he discovered this equation! He had found a fundamental connection between two seemingly unrelated phenomena: electromagnetic fields and light.

? Exercise 7.3.1

The wave equation was obtained by (1) finding the **E** field produced by the changing **B** field, (2) finding the **B** field produced by the changing **E** field, and combining the two results. Which of Maxwell's equations was the basis of step (1) and which of step (2)?

Answer (step 1)

Faraday’s law

Answer (step 2)

the Ampère-Maxwell law

How the **E** and **B** Fields Are Related

So far, we have seen that the rates of change of different components of the **E** and **B** fields are related, that the electromagnetic wave is transverse, and that the wave propagates at speed **c**. We next show what Maxwell’s equations imply about the ratio of the **E** and **B** field magnitudes and the relative directions of the **E** and **B** fields.

We now consider solutions to Equation 7.3.4 in the form of plane waves for the electric field:

$$E_y(x,t) = E_0 \cos(kx - \omega t). \tag{7.3.10}$$

We have arbitrarily taken the wave to be traveling in the **+x**-direction and chosen its phase so that the maximum field strength occurs at the origin at time $t = 0$. We are justified in considering only sines and cosines in this way, and generalizing the results, because Fourier’s theorem implies we can express any wave, including even square step functions, as a superposition of sines and cosines.

At any one specific point in space, the **E** field oscillates sinusoidally at angular frequency ω between $+E_0$ and $-E_0$ and similarly, the **B** field oscillates between $+B_0$ and $-B_0$. The amplitude of the wave is the maximum value of $E_y(x,t)$. The period of oscillation **T** is the time required for a complete oscillation. The frequency **f** is the number of complete oscillations per unit of time, and is related to the angular frequency ω by $\omega = 2\pi f$. The wavelength λ is the distance covered by one complete cycle of the wave, and the wavenumber **k** is the number of wavelengths that fit into a distance of 2π in the units being used. These quantities are related in the same way as for a mechanical wave:

$$\omega = 2\pi f,\ f = \frac{1}{T},\ k = \frac{2\pi}{\lambda},\ and\ c = f\lambda = \omega/k.$$

Given that the solution of E_y has the form shown in Equation ???, we need to determine the B field that accompanies it. From Equation 7.3.11, the magnetic field component B_z must obey

$$\frac{\partial B_z}{\partial t} = -\frac{\partial E_y}{\partial x}$$

$$\frac{\partial B_z}{\partial t} = -\frac{\partial}{\partial x}E_0 \cos(kx - \omega t) = kE_0\, sin\,(kx - \omega t). \tag{7.3.11}$$

Because the solution for the **B**-field pattern of the wave propagates in the +**x**-direction at the same speed **c** as the **E-**field pattern, it must be a function of $k(x - ct) = kx - \omega t$. Thus, we conclude from Equation 7.3.8 that B_z is

$$B_z(x,t) = \frac{k}{\omega}E_0 \cos(kx - \omega t) = \frac{1}{c}E_0 \cos(kx - \omega t).$$

These results may be written as

$$E_y(x,t) = E_0 \cos(kx - \omega t)$$

$$B_z(x,t) = B_0 \cos(kx - \omega t) \tag{7.3.12}$$

$$\frac{E_y}{B_z} = \frac{E_0}{B_0} = c. \tag{7.3.13}$$

Therefore, the peaks of the **E** and **B** fields coincide, as do the troughs of the wave, and at each point, the **E** and **B** fields are in the same ratio equal to the speed of light **c**. The plane wave has the form shown in Figure 7.3.3.

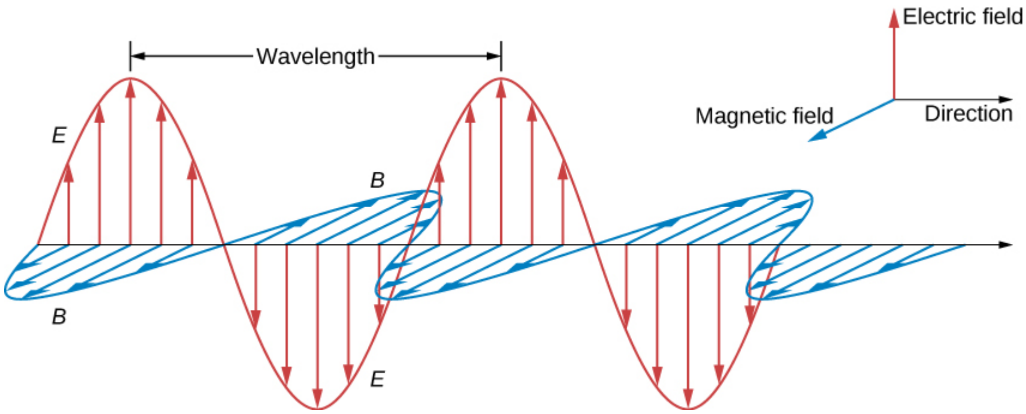


Figure 7.3.3: The plane wave solution of Maxwell's equations has the B field directly proportional to the E field at each point, with the relative directions shown.

Example 7.3.1: Calculating B-Field Strength in an Electromagnetic Wave

What is the maximum strength of the **B** field in an electromagnetic wave that has a maximum **E**-field strength of 1000 V/m?

Strategy

To find the **B**-field strength, we rearrange Equation 7.3.10 to solve for B, yielding

$$B = \frac{E}{c}.$$

Solution We are given **E**, and **c** is the speed of light. Entering these into the expression for **B** yields

$$B = \frac{1000\ V/m}{3.00 \times 10^8\ m/s} = 3.33 \times 10^{-6} T.$$

Significance

The **B**-field strength is less than a tenth of Earth's admittedly weak magnetic field. This means that a relatively strong electric field of 1000 V/m is accompanied by a relatively weak magnetic field.

Changing electric fields create relatively weak magnetic fields. The combined electric and magnetic fields can be detected in electromagnetic waves, however, by taking advantage of the phenomenon of resonance, as Hertz did. A system with the same natural frequency as the electromagnetic wave can be made to oscillate. All radio and TV receivers use this principle to pick up and then amplify weak electromagnetic waves, while rejecting all others not at their resonant frequency.

Exercise 7.3.2

What conclusions did our analysis of Maxwell's equations lead to about these properties of a plane electromagnetic wave:

a. the relative directions of wave propagation, of the **E** field, and of **B** field,
b. the speed of travel of the wave and how the speed depends on frequency, and
c. the relative magnitudes of the **E** and **B** fields.

Answer a

The directions of wave propagation, of the **E** field, and of **B** field are all mutually perpendicular.

Answer b

The speed of the electromagnetic wave is the speed of light $c = 1/\sqrt{\epsilon_0 \mu_0}$ independent of frequency.

Answer c

The ratio of electric and magnetic field amplitudes is $E/B = c$.

Production and Detection of Electromagnetic Waves

A steady electric current produces a magnetic field that is constant in time and which does not propagate as a wave. Accelerating charges, however, produce electromagnetic waves. An electric charge oscillating up and down, or an alternating current or flow of charge in a conductor, emit radiation at the frequencies of their oscillations. The electromagnetic field of a **dipole antenna** is shown in Figure $7.3.4$. The positive and negative charges on the two conductors are made to reverse at the desired frequency by the output of a transmitter as the power source. The continually changing current accelerates charge in the antenna, and this results in an oscillating electric field a distance away from the antenna. The changing electric fields produce changing magnetic fields that in turn produce changing electric fields, which thereby propagate as electromagnetic waves. The frequency of this radiation is the same as the frequency of the ac source that is accelerating the electrons in the antenna. The two conducting elements of the dipole antenna are commonly straight wires. The total length of the two wires is typically about one-half of the desired wavelength (hence, the alternative name **half-wave antenna**), because this allows standing waves to be set up and enhances the effectiveness of the radiation.

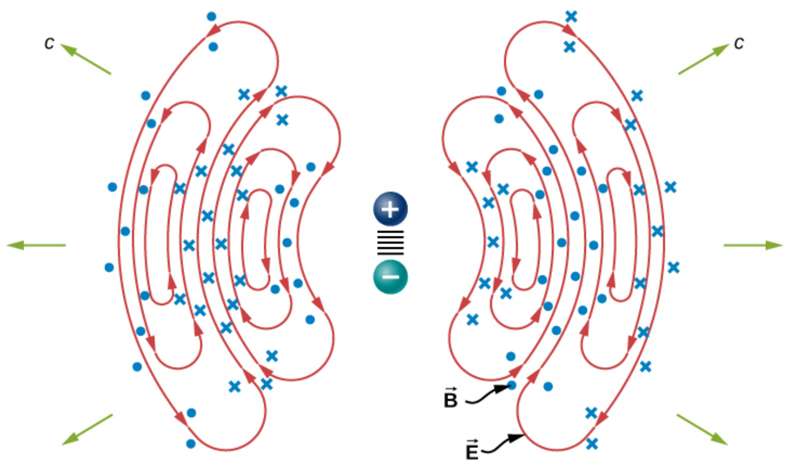


Figure $7.3.4$: The oscillatory motion of the charges in a dipole antenna produces electromagnetic radiation.

The electric field lines in one plane are shown. The magnetic field is perpendicular to this plane. This radiation field has cylindrical symmetry around the axis of the dipole. Field lines near the dipole are not shown. The pattern is not at all uniform in all directions. The strongest signal is in directions perpendicular to the axis of the antenna, which would be horizontal if the antenna is mounted vertically. There is zero intensity along the axis of the antenna. The fields detected far from the antenna are from the changing electric and magnetic fields inducing each other and traveling as electromagnetic waves. Far from the antenna, the wave fronts, or surfaces of equal phase for the electromagnetic wave, are almost spherical. Even farther from the antenna, the radiation propagates like electromagnetic plane waves.

The electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving electromagnetic signals works in reverse. Incoming electromagnetic waves induce oscillating currents in the antenna, each at its own frequency. The radio receiver includes a tuner circuit, whose resonant frequency can be adjusted. The tuner responds strongly to tshe desired frequency but not others, allowing the user to tune to the desired broadcast. Electrical components amplify the signal formed by the moving electrons. The signal is then converted into an audio and/or video format.

Note

Use this simulation to broadcast radio waves. Wiggle the transmitter electron manually or have it oscillate automatically. Display the field as a curve or vectors. The strip chart shows the electron positions at the transmitter and at the receiver.

- **16.3: Plane Electromagnetic Waves** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/university-physics-volume-2.

7.4: Energy Carried by Electromagnetic Waves

Learning Objectives

By the end of this section, you will be able to:

- Express the time-averaged energy density of electromagnetic waves in terms of their electric and magnetic field amplitudes
- Calculate the Poynting vector and the energy intensity of electromagnetic waves
- Explain how the energy of an electromagnetic wave depends on its amplitude, whereas the energy of a photon is proportional to its frequency

Anyone who has used a microwave oven knows there is energy in electromagnetic waves. Sometimes this energy is obvious, such as in the warmth of the summer Sun. Other times, it is subtle, such as the unfelt energy of gamma rays, which can destroy living cells.

Electromagnetic waves bring energy into a system by virtue of their electric and magnetic fields. These fields can exert forces and move charges in the system and, thus, do work on them. However, there is energy in an electromagnetic wave itself, whether it is absorbed or not. Once created, the fields carry energy away from a source. If some energy is later absorbed, the field strengths are diminished and anything left travels on.

Clearly, the larger the strength of the electric and magnetic fields, the more work they can do and the greater the energy the electromagnetic wave carries. In electromagnetic waves, the amplitude is the maximum field strength of the electric and magnetic fields (Figure $7.4.1$). The wave energy is determined by the wave amplitude.

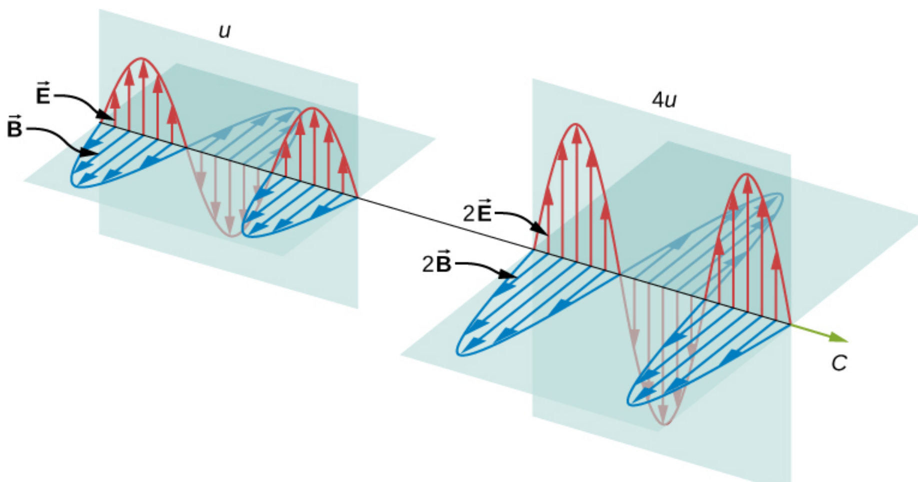


Figure $7.4.1$: Energy carried by a wave depends on its amplitude. With electromagnetic waves, doubling the **E** fields and **B** fields quadruples the energy density **u** and the energy flux **uc**.

For a plane wave traveling in the direction of the positive **x**-axis with the phase of the wave chosen so that the wave maximum is at the origin at $t=0$, the electric and magnetic fields obey the equations

$$E_y(x,t)=E_0\,\cos(kx-\omega t)$$

$$B_x(x,t)=B_0\,\cos(kx-\omega t).$$

The energy in any part of the electromagnetic wave is the sum of the energies of the electric and magnetic fields. This energy per unit volume, or energy density **u**, is the sum of the energy density from the electric field and the energy density from the magnetic field. Expressions for both field energy densities were discussed earlier (u_E in Capacitance and u_B in Inductance). Combining these the contributions, we obtain

$$u(x,t)=u_E+u_B=\frac{1}{2}\epsilon_0E^2+\frac{1}{2\mu_0}B^2.$$

The expression $E = cB = \frac{1}{\sqrt{\epsilon_0 \mu_0}} B$ then shows that the magnetic energy density u_B and electric energy density u_E are equal, despite the fact that changing electric fields generally produce only small magnetic fields. The equality of the electric and magnetic energy densities leads to

$$u(x,t) = \epsilon_0 E^2 = \frac{B^2}{\mu_0}. \tag{7.4.1}$$

The energy density moves with the electric and magnetic fields in a similar manner to the waves themselves.

We can find the rate of transport of energy by considering a small time interval Δt. As shown in Figure $7.4.2$, the energy contained in a cylinder of length $c\Delta t$ and cross-sectional area **A** passes through the cross-sectional plane in the interval Δt.

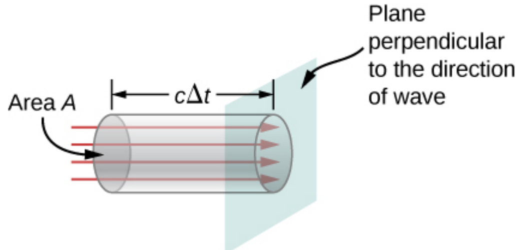


Figure $7.4.2$: The energy $uAc\Delta t$ contained in the electric and magnetic fields of the electromagnetic wave in the volume $Ac\Delta t$ passes through the area A in time Δt.

The energy passing through area A in time Δt is

$$u \times volume = uAc\Delta t.$$

The energy per unit area per unit time passing through a plane perpendicular to the wave, called the **energy flux** and denoted by S, can be calculated by dividing the energy by the area A and the time interval Δt.

$$S = \frac{\text{Energy passing area } A \text{ in time } \Delta t}{A\Delta t} = uc = \epsilon_0 c E^2 = \frac{1}{\mu_0} EB.$$

More generally, the flux of energy through any surface also depends on the orientation of the surface. To take the direction into account, we introduce a vector $\vec{S}$, called the **Poynting vector**, with the following definition:

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}.$$

The cross-product of $\vec{E}$ and $\vec{B}$ points in the direction perpendicular to both vectors. To confirm that the direction of $\vec{S}$ is that of wave propagation, and not its negative, return to Figure 16.3.2. Note that Lenz's and Faraday's laws imply that when the magnetic field shown is increasing in time, the electric field is greater at x than at $x + \Delta x$. The electric field is decreasing with increasing x at the given time and location. The proportionality between electric and magnetic fields requires the electric field to increase in time along with the magnetic field. This is possible only if the wave is propagating to the right in the diagram, in which case, the relative orientations show that $\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}$ is specifically in the direction of propagation of the electromagnetic wave.

The energy flux at any place also varies in time, as can be seen by substituting u from Equation 16.3.19 into Equation 7.4.1.

$$S(x,t) = c\epsilon_0 E_0^2 \cos^2(kx - \omega t) \tag{7.4.2}$$

Because the frequency of visible light is very high, of the order of $10^{14}\ Hz$, the energy flux for visible light through any area is an extremely rapidly varying quantity. Most measuring devices, including our eyes, detect only an average over many cycles. The time average of the energy flux is the **intensity** I of the electromagnetic wave and is the power per unit area. It can be expressed by averaging the cosine function in Equation 7.4.2 over one complete cycle, which is the same as time-averaging over many cycles (here, T is one period):

$$I = S_{avg} = c\epsilon_0 E_0^2 \frac{1}{T} \int_0^T \cos^2\left(2\pi \frac{t}{T}\right) dt. \tag{7.4.3}$$

We can either evaluate the integral, or else note that because the sine and cosine differ merely in phase, the average over a complete cycle for $cos^2(\xi)$ is the same as for $sin^2(\xi)$, to obtain

$$\langle\cos^2\xi\rangle = \frac{1}{2}[\langle\cos^2\xi\rangle + \langle\sin^2\xi\rangle] = \frac{1}{2}\langle 1\rangle = \frac{1}{2}.$$

where the angle brackets $\langle\ldots\rangle$ stand for the time-averaging operation. The intensity of light moving at speed c in vacuum is then found to be

$$I = S_{avg} = \frac{1}{2}c\epsilon_0 E_0^2 \tag{7.4.4}$$

in terms of the maximum electric field strength E_0, which is also the electric field amplitude. Algebraic manipulation produces the relationship

$$I = \frac{cB_0^2}{2\mu_0} \tag{7.4.5}$$

where B_0 is the magnetic field amplitude, which is the same as the maximum magnetic field strength. One more expression for I_{avg} in terms of both electric and magnetic field strengths is useful. Substituting the fact that $cB_0 = E_0$, the previous expression becomes

$$I = \frac{E_0 B_0}{2\mu_0}. \tag{7.4.6}$$

We can use whichever of the three preceding equations is most convenient, because the three equations are really just different versions of the same result: The energy in a wave is related to amplitude squared. Furthermore, because these equations are based on the assumption that the electromagnetic waves are sinusoidal, the peak intensity is twice the average intensity; that is, $I_0 = 2I$.

Example 7.4.1: A Laser Beam

The beam from a small laboratory laser typically has an intensity of about $1.0\times 10^{-3} W/m^2$. Assuming that the beam is composed of plane waves, calculate the amplitudes of the electric and magnetic fields in the beam.

Strategy

Use the equation expressing intensity in terms of electric field to calculate the electric field from the intensity.

Solution

From Equation 7.4.4, the intensity of the laser beam is

$$I = \frac{1}{2}c\epsilon_0 E_0^2.$$

The amplitude of the electric field is therefore

$$\begin{aligned} E_0 &= \sqrt{\frac{2}{c\epsilon_0}I} \\ &= \sqrt{\frac{2}{(3.00\times 10^8 m/s)(8.85\times 10^{-12} F/m)}(1.0\times 10^{-3} W/m^2)} \\ &= 0.87\, V/m. \end{aligned}$$

The amplitude of the magnetic field can be obtained from:

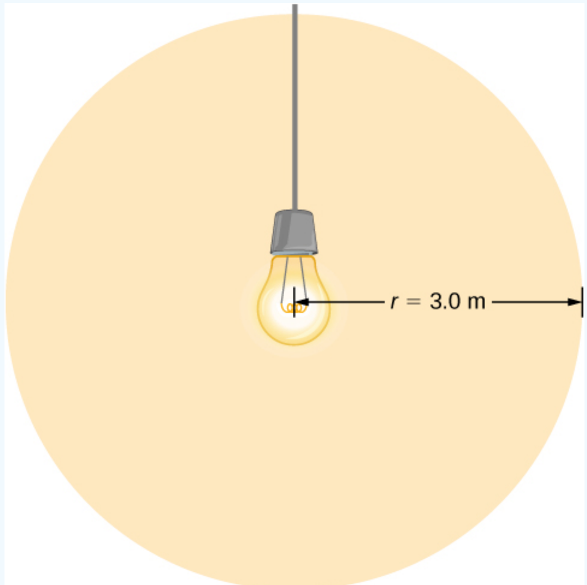
$$B_0 = \frac{E_0}{c} = 2.9\times 10^{-9}\, T.$$

Light Bulb Fields

A light bulb emits 5.00 W of power as visible light. What are the average electric and magnetic fields from the light at a distance of 3.0 m?

Strategy

Assume the bulb’s power output **P** is distributed uniformly over a sphere of radius 3.0 m to calculate the intensity, and from it, the electric field.



Solution

The power radiated as visible light is then

$$I = \frac{P}{4\pi r^2} = \frac{c\epsilon_0 E_0^2}{2},$$

$$E_0 = \sqrt{2\frac{P}{4\pi r^2 c\epsilon_0}} = \sqrt{2\frac{5.00\,W}{4\pi(3.0\,m)^2(3.00\times10^8\,m/s)(8.85\times10^{-12}\,C^2/N\cdot m^2)}} = 5.77\,N/C,$$

$$B_0 = E_0/c = 1.92\times10^{-8}\,T\,.$$

Significance

The intensity **I** falls off as the distance squared if the radiation is dispersed uniformly in all directions.

Radio Range

A 60-kW radio transmitter on Earth sends its signal to a satellite 100 km away (Figure 7.4.3). At what distance in the same direction would the signal have the same maximum field strength if the transmitter’s output power were increased to 90 kW?

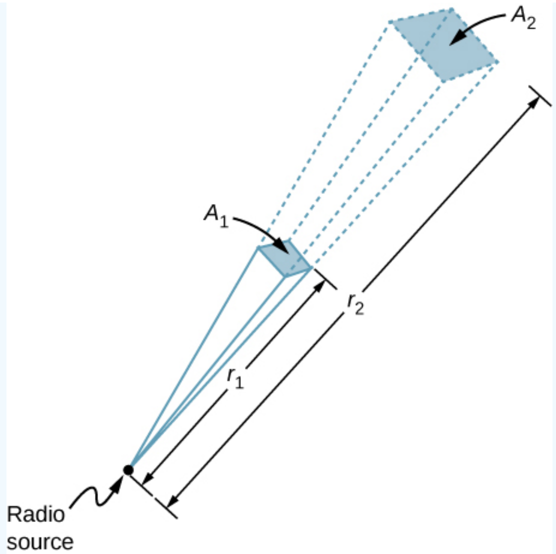


Figure 7.4.3: In three dimensions, a signal spreads over a solid angle as it travels outward from its source.

Strategy

The area over which the power in a particular direction is dispersed increases as distance squared, as illustrated in Figure 7.4.3. Change the power output **P** by a factor of (90 kW/60 kW) and change the area by the same factor to keep $I = \frac{P}{A} = \frac{c\epsilon_0 E_0^2}{2}$ the same. Then use the proportion of area **A** in the diagram to distance squared to find the distance that produces the calculated change in area.

Solution

Using the proportionality of the areas to the squares of the distances, and solving, we obtain from the diagram

$$\frac{r_2^2}{r_1^2} = \frac{A_2}{A_1} = \frac{90\,W}{60\,W},$$

$$r_2 = \sqrt{\frac{90}{60}}(100\,km)$$

$$= 122\,km.$$

Significance

The range of a radio signal is the maximum distance between the transmitter and receiver that allows for normal operation. In the absence of complications such as reflections from obstacles, the intensity follows an inverse square law, and doubling the range would require multiplying the power by four.

7.5: The Electromagnetic Spectrum

Learning Objectives

By the end of this section, you will be able to:

- Explain how electromagnetic waves are divided into different ranges, depending on wavelength and corresponding frequency
- Describe how electromagnetic waves in different categories are produced
- Describe some of the many practical everyday applications of electromagnetic waves

Electromagnetic waves have a vast range of practical everyday applications that includes such diverse uses as communication by cell phone and radio broadcasting, WiFi, cooking, vision, medical imaging, and treating cancer. In this module, we discuss how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on. We also summarize some of the main applications for each range.

The different categories of electromagnetic waves differ in their wavelength range, or equivalently, in their corresponding frequency ranges. Their properties change smoothly from one frequency range to the next, with different applications in each range. A brief overview of the production and utilization of electromagnetic waves is found in Table $7.5.1$.

Table $7.5.1$: Electromagnetic Waves

Type of wave	Production	Applications	Issues
Radio	Accelerating charges	Communications, Remote controls, MRI	Requires control for band use
Microwaves	Accelerating charges and thermal agitation	Communications, Ovens, Radar, Cell phone use	
Infrared	Thermal agitation and electronic transitions	Thermal imaging, Heating	Absorbed by atmosphere, Greenhouse effect
Visible light	Thermal agitation and electronic transitions	Photosynthesis, Human vision	
Ultraviolet	Thermal agitation and electronic transitions	Sterilization, Vitamin D production	Ozone depletion, Cancer causing
X-rays	Inner electronic transitions and fast collisions	Security, Medical diagnosis, Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicine, Security, Medical diagnosis, Cancer therapy	Cancer causing, Radiation damage

The relationship $c = f\lambda$ between frequency **f** and wavelength λ applies to all waves and ensures that greater frequency means smaller wavelength. Figure $7.5.2$ shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies - that is, it shows the electromagnetic spectrum.

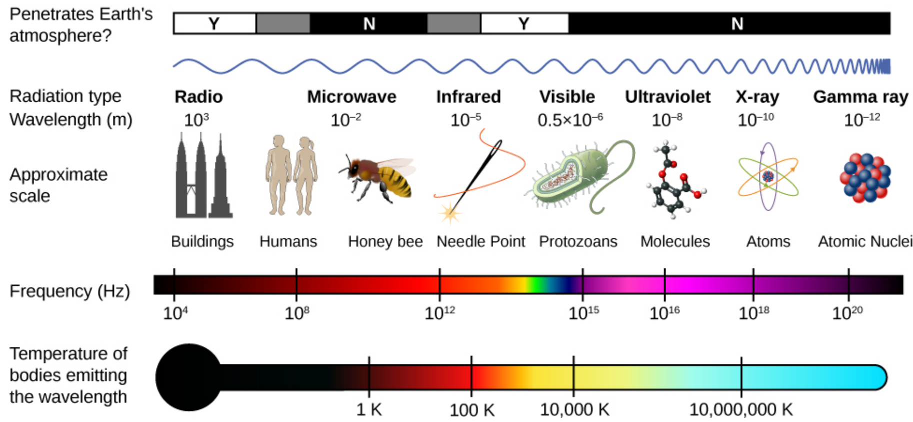


Figure $7.5.1$: The electromagnetic spectrum, showing the major categories of electromagnetic waves.

Radio Waves

The term **radio waves** refers to electromagnetic radiation with wavelengths greater than about 0.1 m. Radio waves are commonly used for audio communications (i.e., for radios), but the term is used for electromagnetic waves in this range regardless of their application. Radio waves typically result from an alternating current in the wires of a broadcast antenna. They cover a very broad wavelength range and are divided into many subranges, including microwaves, electromagnetic waves used for AM and FM radio, cellular telephones, and TV signals.

There is no lowest frequency of radio waves, but ELF waves, or "extremely low frequency" are among the lowest frequencies commonly encountered, from 3 Hz to 3 kHz. The accelerating charge in the ac currents of electrical power lines produce electromagnetic waves in this range. ELF waves are able to penetrate sea water, which strongly absorbs electromagnetic waves of higher frequency, and therefore are useful for submarine communications.

In order to use an electromagnetic wave to transmit information, the amplitude, frequency, or phase of the wave is **modulated**, or varied in a controlled way that encodes the intended information into the wave. In AM radio transmission, the amplitude of the wave is modulated to mimic the vibrations of the sound being conveyed. Fourier's theorem implies that the modulated AM wave amounts to a superposition of waves covering some narrow frequency range. Each AM station is assigned a specific carrier frequency that, by international agreement, is allowed to vary by $\pm 5\,kHz$. In FM radio transmission, the frequency of the wave is modulated to carry this information, as illustrated in Figure $7.5.2$, and the frequency of each station is allowed to use 100 kHz on each side of its carrier frequency. The electromagnetic wave produces a current in a receiving antenna, and the radio or television processes the signal to produce the sound and any image. The higher the frequency of the radio wave used to carry the data, the greater the detailed variation of the wave that can be carried by modulating it over each time unit, and the more data that can be transmitted per unit of time. The assigned frequencies for AM broadcasting are 540 to 1600 kHz, and for FM are 88 MHz to108 MHz.

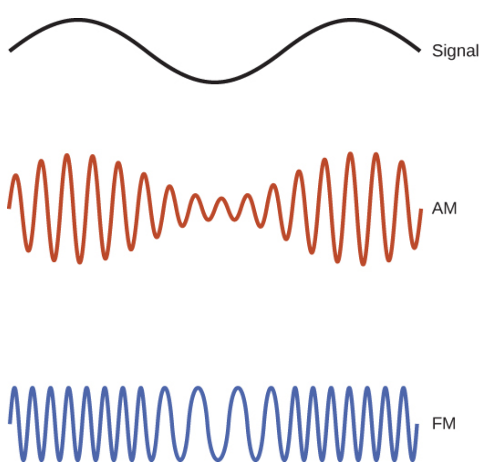


Figure $7.5.2$: Electromagnetic waves are used to carry communications signals by varying the wave's amplitude (AM), its frequency (FM), or its phase.

Cell phone conversations, and **television** voice and video images are commonly transmitted as digital data, by converting the signal into a sequence of binary ones and zeros. This allows clearer data transmission when the signal is weak, and allows using computer algorithms to compress the digital data to transmit more data in each frequency range. Computer data as well is transmitted as a sequence of binary ones and zeros, each one or zero constituting one bit of data.

Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about $10^9 Hz$ to nearly $10^{12} Hz$. Their high frequencies correspond to short wavelengths compared with other radio waves—hence the name "microwave." Microwaves also occur naturally as the cosmic background radiation left over from the origin of the universe. Along with other ranges of electromagnetic waves, they are part of the radiation that any object above absolute zero emits and absorbs because of **thermal agitation**, that is, from the thermal motion of its atoms and molecules.

Most satellite-transmitted information is carried on **microwaves**. **Radar** is a common application of microwaves. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds, aircraft, or even the surface of Venus.

Microwaves of 2.45 GHz are commonly used in microwave ovens. The electrons in a water molecule tend to remain closer to the oxygen nucleus than the hydrogen nuclei (Figure $7.5.3$). This creates two separated centers of equal and opposite charges, giving the molecule a dipole moment. The oscillating electric field of the microwaves inside the oven exerts a torque that tends to align each molecule first in one direction and then in the other, with the motion of each molecule coupled to others around it. This pumps energy into the continual thermal motion of the water to heat the food. The plate under the food contains no water, and remains relatively unheated.

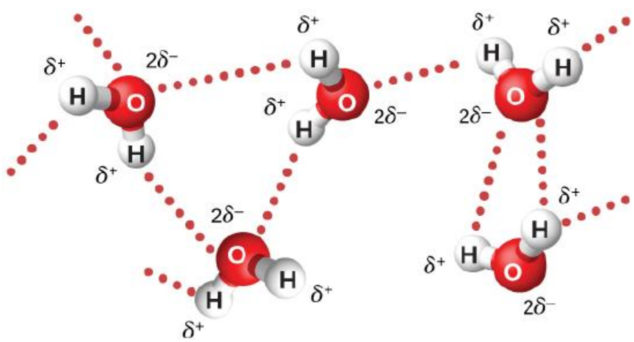


Figure $7.5.3$: The oscillating electric field in a microwave oven exerts a torque on water molecules because of their dipole moment, and the torque reverses direction 4.90×10^9 times per second. Interactions between the molecules distributes the energy being pumped into them. The δ^+ and δ^- denote the charge distribution on the molecules.

The microwaves in a microwave oven reflect off the walls of the oven, so that the superposition of waves produces standing waves, similar to the standing waves of a vibrating guitar or violin string (Normal Modes of a Standing Sound Wave). A rotating fan acts as a stirrer by reflecting the microwaves in different directions, and food turntables, help spread out the hot spots.

Example $7.5.1$: Why Microwave Ovens Heat Unevenly

How far apart are the hotspots in a 2.45-GHz microwave oven?

Strategy

Consider the waves along one direction in the oven, being reflected at the opposite wall from where they are generated.

Solution

The antinodes, where maximum intensity occurs, are half the wavelength apart, with separation

$$d = \frac{1}{2}\lambda \tag{7.5.1}$$

$$= \frac{1}{2}\frac{c}{f} \tag{7.5.2}$$

$$= \frac{3.00 \times 10^8 m/s}{2(2.45 \times 10^9\ Hz)} \tag{7.5.3}$$

$$= 6.02\,cm. \tag{7.5.4}$$

Significance

The distance between the hot spots in a microwave oven are determined by the wavelength of the microwaves.

A cell phone has a radio receiver and a weak radio transmitter, both of which can quickly tune to hundreds of specifically assigned microwave frequencies. The low intensity of the transmitted signal gives it an intentionally limited range. A ground-based system links the phone to only to the broadcast tower assigned to the specific small area, or cell, and smoothly transitions its connection to the next cell when the signal reception there is the stronger one. This enables a cell phone to be used while changing location.

Microwaves also provide the **WiFi** that enables owners of cell phones, laptop computers, and similar devices to connect wirelessly to the Internet at home and at coffee shops and airports. A wireless WiFi router is a device that exchanges data over the Internet through the cable or another connection, and uses microwaves to exchange the data wirelessly with devices such as cell phones and computers. The term WiFi itself refers to the standards followed in modulating and analyzing the microwaves so that wireless routers and devices from different manufacturers work compatibly with one another. The computer data in each direction consist of sequences of binary zeros and ones, each corresponding to a binary bit. The microwaves are in the range of 2.4 GHz to 5.0 GHz range.

Other wireless technologies also use microwaves to provide everyday communications between devices. **Bluetooth** developed alongside WiFi as a standard for radio communication in the 2.4-GHz range between nearby devices, for example, to link to headphones and audio earpieces to devices such as radios, or a driver's cell phone to a hands-free device to allow answering phone calls without fumbling directly with the cell phone.

Microwaves find use also in radio tagging, using RFID (radio frequency identification) technology. Examples are RFID tags attached to store merchandize, transponder for toll booths use attached to the windshield of a car, or even a chip embedded into a pet's skin. The device responds to a microwave signal by emitting a signal of its own with encoded information, allowing stores to quickly ring up items at their cash registers, drivers to charge tolls to their account without stopping, and lost pets to be reunited with their owners. NFC (near field communication) works similarly, except it is much shorter range. Its mechanism of interaction is the induced magnetic field at microwave frequencies between two coils. Cell phones that have NFC capability and the right software can supply information for purchases using the cell phone instead of a physical credit card. The very short range of the data transfer is a desired security feature in this case.

Infrared Radiation

The boundary between the microwave and infrared regions of the electromagnetic spectrum is not well defined (Figure $7.5.1$). **Infrared radiation** is generally produced by thermal motion, and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce **infrared radiation**. About half of the solar energy arriving at Earth is in the infrared region, with most of the rest in the visible part of the spectrum. About 23% of the solar energy is absorbed in the atmosphere, about 48% is absorbed at Earth's surface, and about 29% is reflected back into space.

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means "below red." Water molecules rotate and vibrate particularly well at infrared frequencies. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, including those called **quartz heaters**, to preferentially warm us because we absorb infrared better than our surroundings.

The familiar handheld "remotes" for changing channels and settings on television sets often transmit their signal by modulating an infrared beam. If you try to use a TV remote without the infrared emitter being in direct line of sight with the infrared detector, you may find the television not responding. Some remotes use Bluetooth instead and reduce this annoyance.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum between about 400 nm and about 750 nm to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions.

Red light has the lowest frequencies and longest wavelengths, whereas violet has the highest frequencies and shortest wavelengths (Figure $7.5.4$). Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the sun yellowish in appearance.

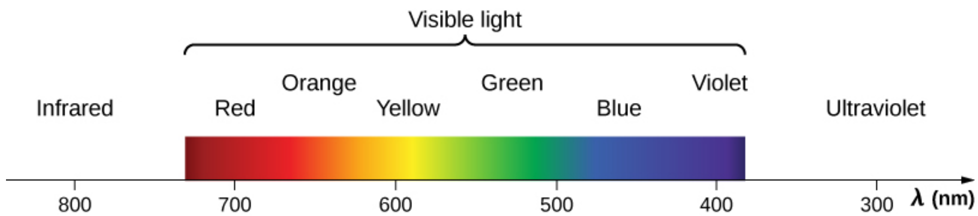


Figure $7.5.4$. A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly distinct, nor are those between the seven rainbow colors.

Living things - plants and animals - have evolved to utilize and respond to parts of the electromagnetic spectrum in which they are embedded. We enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis uses parts of the visible spectrum to make sugars.

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. The highest-frequency ultraviolet overlaps with the lowest-frequency X-rays. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies. Ultraviolet is produced by atomic and molecular motions and electronic transitions.

UV radiation from the Sun is broadly subdivided into three wavelength ranges: UV-A (320–400 nm) is the lowest frequency, then UV-B (290–320 nm) and UV-C (220–290 nm). Most UV-B and all UV-C are absorbed by ozone (O_3) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching Earth’s surface is UV-A.

Sunburn is caused by large exposures to UV-B and UV-C, and repeated exposure can increase the likelihood of skin cancer. The tanning response is a defense mechanism in which the body produces pigments in inert skin layers to reduce exposure of the living cells below.

As examined in a later chapter, the shorter the wavelength of light, the greater the energy change of an atom or molecule that absorbs the light in an electronic transition. This makes short-wavelength ultraviolet light damaging to living cells. It also explains why ultraviolet radiation is better able than visible light to cause some materials to glow, or **fluoresce**.

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin results from exposure to UV-B radiation, generally from sunlight. Several studies suggest vitamin D deficiency is associated with the development of a range of cancers (prostate, breast, colon), as well as osteoporosis. Low-intensity ultraviolet has applications such as providing the energy to cause certain dyes to fluoresce and emit visible light, for example, in printed money to display hidden watermarks as counterfeit protection.

X-Rays

X-rays have wavelengths from about $10^{-8} m$ to $10^{-12} m$. They have shorter wavelengths, and higher frequencies, than ultraviolet, so that the energy they transfer at an atomic level is greater. As a result, X-rays have adverse effects on living cells similar to those of ultraviolet radiation, but they are more penetrating. Cancer and genetic defects can be induced by X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted, and these were designated as alpha, beta, and gamma rays. The most penetrating nuclear radiation, the **gamma ray** γ -ray) was later found to be an extremely high-frequency electromagnetic wave.

The lower end of the γ-ray frequency range overlaps the upper end of the X-ray range. Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. The name “gamma rays” is generally used for electromagnetic radiation emitted by a nucleus, while X-rays are generally produced by bombarding a target with energetic electrons in an X-ray tube. At higher frequencies, γ-rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

Use this simulation to explore how light interacts with molecules in our atmosphere.

- Explore how light interacts with molecules in our atmosphere.
- Identify that absorption of light depends on the molecule and the type of light.
- Relate the energy of the light to the resulting motion.
- Identify that energy increases from microwave to ultraviolet.
- Predict the motion of a molecule based on the type of light it absorbs.

Exercise 7.5.1

How do the electromagnetic waves for the different kinds of electromagnetic radiation differ?

Answer

They fall into different ranges of wavelength, and therefore also different corresponding ranges of frequency.

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Phys1140: Introductory Physics II: Part 1

Webpages: 47

All licenses found:

- CC BY 4.0: 100% (47 pages)

By Page

- Phys1140: Introductory Physics II: Part 1 - *CC BY 4.0*
  - Front Matter - *CC BY 4.0*
    - TitlePage - *CC BY 4.0*
    - InfoPage - *CC BY 4.0*
    - Table of Contents - *CC BY 4.0*
    - Licensing - *CC BY 4.0*
  - 1: Charges and Conductors - *CC BY 4.0*
    - 1.1: Prelude to Electric Charges and Fields - *CC BY 4.0*
    - 1.2: Electric Charge - *CC BY 4.0*
    - 1.3: Conductors, Insulators, and Charging by Induction - *CC BY 4.0*
  - 2: Electric Fields - *CC BY 4.0*
    - 2.1: Coulomb's Law - *CC BY 4.0*
    - 2.2: Electric Field - *CC BY 4.0*
    - 2.3: Calculating Electric Fields of Charge Distributions - *CC BY 4.0*
    - 2.4: Electric Field Lines - *CC BY 4.0*
  - 3: Electric Potentials - *CC BY 4.0*
    - 3.1: Prelude to Electric Potential - *CC BY 4.0*
    - 3.2: Electric Potential Energy - *CC BY 4.0*
    - 3.3: Electric Potential and Potential Difference - *CC BY 4.0*
  - 4: Current and Resistance - *CC BY 4.0*
    - 4.1: Prelude to Current and Resistance - *CC BY 4.0*
    - 4.2: Electrical Current - *CC BY 4.0*
    - 4.3: Model of Conduction in Metals - *CC BY 4.0*
    - 4.4: Resistivity and Resistance - *CC BY 4.0*
    - 4.5: Ohm's Law - *CC BY 4.0*
    - 4.6: Electrical Energy and Power - *CC BY 4.0*
    - 4.7: Superconductors - *CC BY 4.0*
  - 5: Direct Current Circuits - *CC BY 4.0*
    - 5.1: Prelude to Direct-Current Circuits - *CC BY 4.0*
    - 5.2: Electromotive Force - *CC BY 4.0*
    - 5.3: Resistors in Series and Parallel - *CC BY 4.0*
  - 6: Magnetic Fields - *CC BY 4.0*
    - 6.1: Prelude to Magnetic Forces and Fields - *CC BY 4.0*
    - 6.2: Magnetism and Its Historical Discoveries - *CC BY 4.0*
    - 6.3: Magnetic Fields and Lines - *CC BY 4.0*
    - 6.4: Force and Torque on a Current Loop - *CC BY 4.0*
    - 6.5: The Hall Effect - *CC BY 4.0*
  - 7: Electromagnetic Waves - *CC BY 4.0*
    - 7.1: Prelude to Electromagnetic Waves - *CC BY 4.0*
    - 7.2: Maxwell’s Equations and Electromagnetic Waves - *CC BY 4.0*
    - 7.3: Plane Electromagnetic Waves - *CC BY 4.0*
    - 7.4: Energy Carried by Electromagnetic Waves - *CC BY 4.0*
    - 7.5: The Electromagnetic Spectrum - *CC BY 4.0*
  - Back Matter - *CC BY 4.0*
    - Index - *CC BY 4.0*
    - Glossary - *CC BY 4.0*
    - Detailed Licensing - *CC BY 4.0*