CHAPTER 1

The Physics of Electricity

1.1 BASIC QUANTITIES

1.1.1 Introduction

This chapter describes the quantities that are essential to our understanding of electricity: charge, voltage, current, resistance, and electric and magnetic fields. Most students of science and engineering find it very hard to gain an intuitive appreciation of these quantities, since they are not part of the way we normally see and make sense of the world around us. Electrical phenomena have a certain mystique that derives from the difficulty of associating them with our direct experience, but also from the knowledge that they embody a potent, fundamental force of nature.

Electric charge is one of the basic dimensions of physical measurement, along with mass, distance, time and temperature. All other units in physics can be expressed as some combination of these five terms. Unlike the other four, however, charge is more remote from our sensory perception. While we can easily visualize the size of an object, imagine its weight, or anticipate the duration of a process, it is difficult to conceive of "charge" as a tangible phenomenon.

To be sure, electrical processes are vital to our bodies, from cell metabolism to nervous impulses, but we do not usually conceptualize these in terms of electrical quantities or forces. Our most direct and obvious experience of electricity is to receive an electric shock. Here the presence of charge sends such a strong wave of nervous impulses through our body that it produces a distinct and unique sensation. Other firsthand encounters with electricity include hair that defiantly stands on end, a zap from a door knob, and static cling in the laundry. Yet these experiences hardly translate into the context of electric power, where we can witness the *effects* of electricity, such as a glowing light bulb or a rotating motor, while the essential happenings take place silently and concealed within pieces of metal. For the most part, then, electricity remains an abstraction to us, and we rely on numerical and geometric representations—aided by liberal analogies from other areas of the physical world—to form concepts and develop an intuition about it.

Electric Power Systems: A Conceptual Introduction, by Alexandra von Meier

1.1.2 Charge

It was a major scientific accomplishment to integrate an understanding of electricity with fundamental concepts about the microscopic nature of matter. Observations of static electricity like those mentioned earlier were elegantly explained by Benjamin Franklin in the late 1700s as follows: There exist in nature two types of a property called *charge*, arbitrarily labeled "positive" and "negative." Opposite charges attract each other, while like charges repel. When certain materials rub together, one type of charge can be transferred by friction and "charge up" objects that subsequently repel objects of the same kind (hair), or attract objects of a different kind (polyester and cotton, for instance).

Through a host of ingenious experiments,[1] scientists arrived at a model of the atom as being composed of smaller individual particles with opposite charges, held together by their electrical attraction. Specifically, the nucleus of an atom, which constitutes the vast majority of its mass, contains *protons* with a positive charge, and is enshrouded by *electrons* with a negative charge. The nucleus also contains neutrons, which resemble protons, except they have no charge. The electric attraction between protons and electrons just balances the electrons' natural tendency to escape, which results from both their rapid movement, or kinetic energy, and their mutual electric repulsion. (The repulsion among protons in the nucleus is overcome by another type of force called the *strong nuclear interaction*, which only acts over very short distances.)

This model explains both why most materials exhibit no obvious electrical properties, and how they can become "charged" under certain circumstances: The opposite charges carried by electrons and protons are equivalent in magnitude, and when electrons and protons are present in equal numbers (as they are in a normal atom), these charges "cancel" each other in terms of their effect on their environment. Thus, from the outside, the entire atom appears as if it had no charge whatsoever; it is *electrically neutral*.

Yet individual electrons can sometimes escape from their atoms and travel elsewhere. Friction, for instance, can cause electrons to be transferred from one material into another. As a result, the material with excess electrons becomes *negatively charged*, and the material with a deficit of electrons becomes *positively charged* (since the positive charge of its protons is no longer compensated). The ability of electrons to travel also explains the phenomenon of *electric current*, as we will see shortly.

Some atoms or groups of atoms (molecules) naturally occur with a net charge because they contain an imbalanced number of protons and electrons; they are called *ions*. The propensity of an atom or molecule to become an ion—namely, to release electrons or accept additional ones—results from peculiarities in the geometric pattern by which electrons occupy the space around the nuclei. Even electrically neutral molecules can have a local appearance of charge that results from

[1]Almost any introductory physics text will provide examples. For an explanation of the basic concepts of electricity, I recommend Paul Hewitt, *Conceptual Physics*, Tenth Edition (Menlo Park, CA: Addison Wesley, 2006).

imbalances in the spatial distribution of electrons—that is, electrons favoring one side over the other side of the molecule. These electrical phenomena within molecules determine most of the physical and chemical properties of all the substances we know.[2]

While on the microscopic level, one deals with fundamental units of charge (that of a single electron or proton), the practical unit of charge in the context of electric power is the *coulomb* (C). One coulomb corresponds to the charge of 6.25×10^{18} protons. Stated the other way around, one proton has a charge of 1.6×10^{-19} C. One electron has a negative charge of the same magnitude, -1.6×10^{-19} C. In equations, charge is conventionally denoted by the symbol Q or q.

1.1.3 Potential or Voltage

Because like charges repel and opposite charges attract, charge has a natural tendency to "spread out." A local accumulation or deficit of electrons causes a certain "discomfort" or "tension":[3] unless physically restricted, these charges will tend to move in such a way as to relieve the local imbalance. In rigorous physical terms, the discomfort level is expressed as a level of *energy*. This energy (strictly, electrical potential energy), said to be "held" or "possessed" by a charge, is analogous to the mechanical potential energy possessed by a massive object when it is elevated above the ground: we might say that, by virtue of its height, the object has an inherent potential to fall down. A state of lower energy—closer to the ground, or farther away from like charges—represents a more "comfortable" state, with a smaller potential fall.

The potential energy held by an object or charge in a particular location can be specified in two ways that are physically equivalent: first, it is the *work*[4] that would be required in order to move the object or charge *to* that location. For example, it takes work to lift an object; it also takes work to bring an electron near an accumulation of more electrons. Alternatively, the potential energy is the work the object or charge would do in order to move *from* that location, through interacting with the objects in its way. For example, a weight suspended by a rubber band will stretch the rubber band in order to move downward with the pull of gravity (from higher to lower gravitational potential). A charge moving toward a more comfortable location might do work by producing heat in the wire through which it flows.

[2]For example, water owes its amazing liquidity and density at room temperature to the electrical attraction among its neutral molecules that results from each molecule being polarized: casually speaking, the electrons prefer to hang out near the oxygen atom as opposed to the hydrogen atoms of H_2O; a chemist would say that oxygen has a greater *electronegativity* than hydrogen. The resulting attraction between these polarized ends of molecules is called a *hydrogen bond*, which is essential to all aspects of our physical life.

[3]The term *tension* is actually synonymous with voltage or potential, mainly in British usage.

[4]In physics, work is equivalent to and measured in the same units as energy, with the implied sense of exerting a force to "push" or "pull" something over some distance (Work = Force × Distance).

This notion of work is crucial because, as we will see later, it represents the physical basis of transferring and utilizing electrical energy. In order to make this "work" a useful and unambiguous measure, some proper definitions are necessary. The first is to explicitly distinguish the contributions of charge and potential to the total amount of work or energy transferred. Clearly, the amount of work in either direction (higher or lower potential) depends on the amount of mass or charge involved. For example, a heavy weight would stretch a rubber band farther, or even break it. Similarly, a greater charge will do more work in order to move to a lower potential.

On the other hand, we also wish to characterize the location proper, independent of the object or charge there. Thus, we establish the rigorous definition of the electric *potential*, which is synonymous with *voltage* (but more formal). The electric potential is the potential energy possessed by a charge at the location in question, relative to a reference location, divided by the amount of its charge. Casually speaking, we might say that the potential represents a measure of how comfortable or uncomfortable it would be for any charge to reside at that location. A potential or voltage can be positive or negative. A positive voltage implies that a positive charge would be repelled, whereas a negative charge would be attracted to the location; a negative voltage implies the opposite.

Furthermore, we must be careful to specify the "reference" location: namely, the place where the object or charge was moved from or to. In the mechanical context, we specify the height *above ground level*. In electricity, we refer to an electrically neutral place, real or abstract, with *zero* or *ground potential*. Theoretically, one might imagine a place where no other charges are present to exert any forces; in practice, ground potential is any place where positive and negative charges are balanced and their influences cancel. When describing the potential at a single location, it is implicitly the potential *difference* between this and the neutral location. However, potential can also be specified as a difference between two locations of which neither is neutral, like a difference in height.

Because electric potential or voltage equals energy per charge, the units of voltage are equivalent to units of energy divided by units of charge. These units are *volts* (V). One volt is equivalent to one joule per coulomb, where the joule is a standard unit of work or energy.[5]

Note how the notion of a difference always remains implicit in the measurement of volts. A statement like "this wire is at a voltage of 100 volts" means "this wire is at a voltage of 100 volts *relative to ground*," or "the voltage *difference between the wire and the ground* is 100 volts." By contrast, if we say "the battery has a voltage of 1.5 volts," we mean that "the voltage *difference between the two terminals* of the battery is 1.5 volts." Note that the latter statement does not tell us the potential of either terminal in relation to ground, which depends on the type of battery and whether it is connected to other batteries.

[5]A joule can be expressed as a watt-second; 1 kilowatt-hour = 3.6×10^6 joules, as there are 3600 seconds in an hour.

In equations, voltage is conventionally denoted by E, e, V, or v (in a rare and inelegant instance of using the same letter for both the symbol of the quantity and its unit of measurement).

1.1.4 Ground

The term *ground* has a very important and specific meaning in the context of electric circuits: it is an electrically neutral place, meaning that it has zero voltage or potential, which moreover has the ability to absorb excesses of either positive or negative charge and disperse them so as to *remain* neutral regardless of what might be electrically connected to it. The literal ground outdoors has this ability because the Earth as a whole acts as a vast reservoir of charge and is electrically neutral, and because most soils are sufficiently conductive to allow charge to move away from any local accumulation. The term *earth* is synonymous with ground, especially in British usage. A circuit "ground" is constructed simply by creating a pathway for charge into the earth. In the home, this is often done by attaching a wire to metal water pipes. In power systems, ground wires, capable of carrying large currents if necessary, are specifically dug into the earth.

1.1.5 Conductivity

To understand conductivity, we must return to the microscopic view of matter. In most materials, electrons are bound to their atoms or molecules by the attraction to the protons in the nuclei. We have mentioned how special conditions such as friction can cause electrons to escape. In certain materials, some number of electrons are always free to travel. As a result, the material is able to *conduct* electricity. When a charge (i.e., an excess or deficit of electrons) is applied to one side of such a conducting material, the electrons throughout will realign themselves, spreading out by virtue of their mutual repulsion, and thus conduct the charge to the other side.

For this to happen, an individual electron need not travel very far. We can imagine each electron moving a little to the side, giving its neighbor a repulsive "shove," and this shove propagating through the conducting material like a wave of falling dominoes.

The most important conducting materials in our context are metals. The microscopic structure of metals is such that some electrons are always free to travel throughout a fixed lattice of positive ions (the atomic nuclei surrounded by the remaining, tightly bound electrons).[6] While all metals conduct, their *conductivity* varies quantitatively depending on the ease with which electrons can travel, or the extent to which their movement tends to be hampered by microscopic forces and collisions inside the material.

[6]This property can be understood through the periodic table of the elements, which identifies metals as being those types of atoms with one or a few electrons dwelling alone in more distant locations from the nucleus (orbitals), from where they are easily removed (ionized) so as to become free electrons.

Besides metals, there are other types of material that conduct electricity. One is water, or any other fluid, with dissolved ions (such as salt or minerals). In this case, it is not electrons but entire charged molecules that travel through the fluid to carry a current. Only small concentrations of ions are needed to make water conductive; while pure distilled water does not conduct electricity, normal tap water and rain water conduct all too well.[7]

Some materials, including air, can also become temporarily conductive through *ionization*. In the presence of a very strong potential gradient (defined as an *electric field* in Section 1.5), or intense heat, some electrons are stripped from their molecules and become free to travel. A gas in this state is called a *plasma*. Plasmas exist inside stars, nuclear fusion reactors, or fluorescent lights. More often, though, ionization tends to be local and transient: it occurs along a distinct trail, since ionized molecules incite their neighbors to do the same, and charge flows along this trail until the potential difference (charge imbalance) is neutralized. This is precisely what happens in an electric spark across an air gap, an arc between power lines, or a lightning bolt.[8]

In many engineering situations, it is important to predict just when ionization might occur; namely, how great a potential difference over how short a distance will cause "arcing." For air, this varies according to temperature and especially humidity, as well as the presence of other substances like salt suspended in the air. Exact figures for the *ionizing potential* can be found in engineering tables. For units of conductivity and the relationship to resistance, see Section 1.2.

Finally, some materials can become *superconducting*, generally at very low temperatures. Here, electrons undergo a peculiar energetic transition that allows them to travel with extreme ease, unimpeded by any obstructive forces or collisions. Thus, electrons in the superconducting state do no work on anything in their path, and therefore lose no energy. Some ceramic materials attain superconductivity at a temperature easily sustained by cooling with liquid nitrogen (at minus 319°F).[9] While liquid nitrogen is quite cheap in a research setting, large-scale refrigeration systems aimed at taking advantage of superconductivity in electric power applications are generally considered too formidable in cost to be justified by the savings in electric losses (see Section 1.3). Another conceivable application of superconductivity in power systems is superconducting magnetic energy storage (SMES).

1.1.6 Current

When charge travels through a material, an *electric current* is said to flow. The current is quantified in terms of the number of electrons (or equivalent charge, in

[7] In fact, conductivity is used as an indicator of water purity. Of course, it says nothing about the *kind* of ions present, only the amount.

[8] The ionization trail is visible because, as the electrons return to their normal state, the balance of their energy is released in the form of light.

[9] The first such material to be discovered was yttrium-barium-copper oxide, $YBa_2Cu_3O_7$.

the case of ions) moving past a given point in the material in a certain period of time. In other words, current is a *flow rate* of charge. In this way, electric current is analogous to a flow rate of water (say, in gallons per minute) or natural gas (cubic feet per second).

These analogies are also helpful in remembering the distinction between current and voltage. Voltage would be analogous to a height difference (say, between a water reservoir and the downhill end of a pipe), or to a pressure difference (between two ends of a gas pipeline). Intuitively, voltage is a measure of "how badly the stuff wants to get there," and current is a measure of "how much stuff is actually going."

Current is conventionally denoted by the symbol I or i and is measured in units of *amperes* (A), often called "amps." Since current represents a flow rate of charge, the units of current are equivalent to units of charge divided by units of time. Thus, one ampere equals one coulomb per second.

A subject that often causes confusion is the "direction" in which current flows, though in practice, having an accurate picture of this is not all that important. Most often, the reasons one is concerned with current have to do with the amount of power transferred or the amount of heating of the wires, neither of which depend on direction.

When in doubt, we can always refer back to the fact that opposite charges attract and like charges repel. Thus, a positive charge will be attracted by a negative potential, and hence flow toward it, and vice versa: electrons, which have negative charge, flow toward a positive potential or voltage. In a mathematical sense, negative charge flowing in one direction is equivalent to positive charge flowing the opposite way. Indeed, our practical representation of electric current does not distinguish between these two physical phenomena. For example, the current flowing through a lead–acid battery at various times consists of negative electrons in the terminals and wires, and positive ions in the battery fluid; yet these flows are thought of as the same current.

In circuit analysis, it often becomes necessary to define a direction of current flow, so as to know when to add and when to subtract currents that meet on a section of the circuit. The general convention is to label a current flow as "positive" in the direction from positive toward negative potential (as if a *positive* charge were flowing). Once this labeling has been chosen, all currents in the circuit will be computed as positive or negative so as to be consistent with that requirement (positive currents will always point toward lower potential). However, the convention is arbitrary in that one can define the currents throughout an entire circuit "backwards," and obtain just as "correct" a result. In other words, for purposes of calculation, the quantity "current" need not indicate the actual physical direction of traveling charge.

In the power systems context, the notion of directionality is more complicated (and less revealing) because the physical direction of current flow actually alternates (see "Alternating Current" in Section 3.1). Instead, to capture the relationship between two currents (whether they add or subtract), the concept of *phase*, or relative timing, is used.

As for the speed at which current propagates, it is often said that current travels at the speed of light (186,000 miles per second). While this is not quite accurate (just as the speed of light actually varies in different materials), it is usually sufficient to know that current travels *very fast*. Conceptually, it is important to recognize that what is traveling at this high speed is the *pulse* or *signal* of the current, not the individual electrons. For the current to flow, it is also not necessary for all the electrons to physically depart at one end and arrive at the other end of the conductor. Rather, the electrons inside a metal conductor continually move in a more or less random way, wiggling around in different directions at a speed related to the temperature of the material. They then receive a "shove" in one direction by the *electric field* (see Section 1.5.2). We can imagine this shove propagating by way of the electrical repulsion among electrons: each electron need not travel a long distance, just enough to push its neighbor over a bit, which in turn pushes *its* neighbor, and so on. This chain reaction creates a more orderly motion of charge, as opposed to the usual random motion, and is observed macroscopically as the current. It is the signal to "move over" that propagates at essentially the speed of light.[10]

The question of the propagation speed of electric current only becomes relevant when the distance to be covered is so large that the time it takes for a current pulse to travel from one point to another is significant compared to other timing parameters of the circuit. This can be the case for electric transmission lines that extend over many hundreds of miles.[11] However, we will not deal with this problem explicitly (see Chapter 7, "Power Flow Analysis," for how we treat the concept of time in power systems). A circuit that is sufficiently small so that the speed of current is not an issue is called a *lumped circuit*. Circuits are treated as lumped circuits unless otherwise stated.

1.2 OHM'S LAW

It is intuitive that voltage and current would be somehow related. For example, if the potential difference between two ends of a wire is increased, we would expect a greater current to flow, just like the flow rate of gas through a pipeline increases when a greater pressure difference is applied. For most materials, including metallic conductors, this relationship between voltage and current is linear: as the potential difference between the two ends of the conductor increases, the current through the conductor increases proportionally. This statement is expressed in Ohm's law,

$$V = IR$$

[10]We can draw an analogy with an ocean wave: the water itself moves essentially up and down, and it is the "signal" to move up and down that propagates across the surface, at a speed much faster than the bulk motion of water.

[11]For example, traveling down a 500-mile transmission line at the speed of light takes 2.7 milliseconds. Compared to the rate at which alternating current changes direction (60 times per second, or every 16.7 milliseconds), this corresponds to one-sixth of a cycle, which is not negligible.

where V is the voltage, I is the current, and R is the proportionality constant called the *resistance*.

1.2.1 Resistance

To say that Ohm's law is true for a particular conductor is to say that the resistance of this conductor is, in fact, constant with respect to current and voltage. Certain materials and electronic devices exhibit a nonlinear relationship between current and voltage, that is, their resistance varies depending on the voltage applied. The relationship $V = IR$ will still hold at any given time, but the value of R will be a different one for different values of V and I. These nonlinear devices have specialized applications and will not be discussed in this chapter. Resistance also tends to vary with temperature, though a conductor can still obey Ohm's law at any one temperature.[12] For example, the resistance of a copper wire increases as it heats up. In most operating regimes, these variations are negligible. Generally, in any situation where changes in resistance are significant, this is explicitly mentioned. Thus, whenever one encounters the term "resistance" without further elaboration, it is safe to assume that within the given context, this resistance is a fixed, unchanging property of the object in question.

Resistance depends on an object's material composition as well as its shape. For a wire, resistance increases with length, and decreases with cross-sectional area. Again, the analogy to a gas or water pipe is handy: we know that a pipe will allow a higher flow rate for the same pressure difference if it has a greater diameter, while the flow rate will decrease with the length of the pipe. This is due to *friction* in the pipe, and in fact, an analogous "friction" occurs when an electric current travels through a material.

This friction can be explained by referring to the microscopic movement of electrons or ions, and noting that they interact or collide with other particles in the material as they go. The resulting forces tend to impede the movement of the charge carriers and in effect limit the rate at which they pass. These forces vary for different materials because of the different spatial arrangements of electrons and nuclei, and they determine the material's ability to conduct.

This intrinsic material property, independent of size or shape, is called *resistivity* and is denoted by ρ (the Greek lowercase rho). The actual resistance of an object is given by the resistivity multiplied by the length of the object (l) and divided by its cross-sectional area (A):

$$R = \frac{\rho l}{A}$$

The units of resistance are *ohms*, abbreviated Ω (Greek capital omega). By rearranging Ohm's law, we see that resistance equals voltage divided by current. Units of

[12] If we graph V versus I, Ohm's law requires that the graph be a straight line. With temperature, the slope of this line may change.

resistance are thus equivalent to units of voltage divided by units of current. By definition, one ohm equals one volt per ampere ($\Omega = V/A$).

The units of resistivity are ohm-meters (Ω-m), which can be reconstructed through the preceding formula: when ohm-meters are multiplied by meters (for l) and divided by square meters (for A), the result is simply ohms. Resistivity, which is an intrinsic property of a material, is not to be confused with the *resistance per unit length* (usually of a wire), quoted in units of ohms *per* meter (Ω/m). The latter measure already takes into account the wire diameter; it represents, in effect, the quantity ρ/A. The resistivities of different materials in Ω-m can be found in engineering tables.

1.2.2 Conductance

It is sometimes convenient to refer to the resistive property of a material or object in the inverse, as *conductivity* or *conductance*. Conductivity is the inverse of resistivity and is denoted by σ (Greek lowercase sigma): $\sigma = 1/\rho$. For the case of a simple resistor, conductance is the reciprocal of resistance and is usually denoted by G (sometimes g), where $G = 1/R$.[13] Not without humor, the units of conductance are called *mhos*, and 1 mho $= 1/\Omega$. Another name for the mho is the siemens (S); they are identical units. The conductance is related to the conductivity by

$$G = \frac{\sigma A}{l}$$

and the units of σ are thus mhos/m.

For the special case of an insulator, the conductance is zero and the resistance is infinite. For the special case of a superconductor, the resistance is zero and the conductance is, theoretically, infinite (a truly infinite conductance would imply an infinitely large current, which does not actually occur since its magnitude is eventually constrained by the number of electrons available).

Example

Consider two power extension cords, one with twice the wire diameter of the other. If the cords are of the same length and same material, how do their resistances compare?

Since resistance is inversely proportional to area, the smaller wire will have *four times* the resistance. We can see this through the formula $R = \rho l/A$, where ρ and l are the same for both. Thus, using the subscripts 1 and 2 to refer to the two cords, we can write $R_1/R_2 = A_2/A_1$. The areas are given by the familiar geometry formula, $A = \pi(d/2)^2$ (where $\pi = 3.1415\ldots$), which includes the *square* of the diameter or radius. If the length of either cord were doubled, its resistance would also double.

[13] See Section 3.2.4 for the complex case that includes both resistance and reactance.

To put some numbers to this example, consider a typical 25-ft, 16-gauge extension cord, made of a copper conductor. The cross-sectional area of 16-gauge wire is 1.31 mm^2 (or 1.31×10^{-8} m^2) and the resistivity of copper is $\rho = 1.76 \times 10^{-8}$ Ω-m. The resistance per unit length of 16-gauge copper wire is 0.0134 Ω/m, and a 25-ft length of it has a resistance of 0.102 Ω. By contrast, a 10-gauge copper wire of the same length, which has about twice the diameter, has a resistance of only 0.025 Ω.

Suppose the current in the 25-ft, 16-gauge cord is 5 A. What is the voltage difference between the two ends of each conductor?

The voltage drop in the wire is given by Ohm's law, $V = IR$. Thus, $V = 5\ \text{A} \cdot 0.102\ \Omega = 0.51$ V. Because the voltage drop applies to each of the two conductors in the cord, this means that the line voltage, or difference between the two sides of the electrical outlet, will be diminished by about one volt (say, from 120 to 119 V), as seen by the appliance at the end.

1.2.3 Insulation

Insulating materials are used in electric devices to keep current from flowing where it is not desired. They are simply materials with a sufficiently high resistance (or sufficiently low conductance), also known as *dielectric* materials. Typically, plastics or ceramics are used. When an insulator is functional, its resistance is infinite, or the conductance zero, so that zero current flows through it.

Any insulator has a specific voltage regime within which it can be expected to perform. If the voltage difference between two sides of the insulator becomes too large, its insulating properties may break down due to microscopic changes in the material, where it actually becomes conducting. Generally, the thicker the insulator, the higher the voltage difference it can sustain. However, temperature can also be important; for example, plastic wire insulation may melt if the wire becomes too hot.

The insulators often seen on high-voltage equipment consist of strings of ceramic bells, holding the energized wires away from other components (e.g., transmission towers or transformers). The shape of these bells serves to inhibit the formation of arcs along their surface. The number of bells is roughly proportional to the voltage level, though it also depends on climate. For example, the presence of salt water droplets in coastal air encourages ionization and therefore requires more insulation to prevent arcing.

1.3 CIRCUIT FUNDAMENTALS

1.3.1 Static Charge

A current can only flow as long as a potential difference is sustained; in other words, the flowing charge must be replenished. Therefore, some currents have a very short duration. For example, a lightning bolt lasts only a fraction of a second, until the charge imbalance between the clouds and the ground is neutralized.

When charge accumulates in one place, it is called *static charge*, because it is not moving. The reason charge remains static is that it lacks a conducting pathway that enables it to flow toward its opposite charge. When we receive a shock from static electricity—for example, by touching a doorknob—our body is providing just such a pathway. In this example, our body is charged through friction, often on a synthetic carpet, and this charge returns to the ground via the doorknob (the carpet only gives off electrons by rubbing, but does not allow them to flow back). As our fingers approach the doorknob, the air in between is actually ionized momentarily, producing a tiny arc that causes the painful sensation.[14] Static electricity occurs mostly in dry weather, since moisture on the surface of objects makes them sufficiently conductive to prevent accumulations of charge.

However startling and uncomfortable, static electricity encountered in everyday situations is harmless because the amount of charge available is so small,[15] and it is not being replenished. This is true despite the fact that very high voltages can be involved (recall that voltage is energy *per* charge), but these voltages drop instantaneously as soon as the contact is made.

1.3.2 Electric Circuits

In order to produce a sustained flow of current, the potential difference must be maintained. This is achieved by providing a pathway to "recycle" charge to its origin, and a mechanism (called an *electromotive force*, or *emf*[16]) that compels the charge to return to the less "comfortable" potential. Such a setup constitutes an *electric circuit*.

A simple example is a battery connected with two wires to a light bulb. The chemical forces inside the battery do work on the charge to move it to the terminals, where an electric potential is produced and sustained. Specifically, electrons are moved to the negative terminal, and positive ions are moved to the positive electrode, where they produce a deficit of electrons in the positive battery terminal. The wires then provide a path for electrons to flow from the negative to the positive terminal. Because the positive potential is so attractive, these electrons even do work by flowing through the resistive light bulb, causing it to heat up and glow. As soon as the electrons arrive at the positive terminal, they are "lifted" again to the negative potential, allowing the current to continue flowing. In analogy with flowing water, the wires are like pipes that carry water downhill, and the battery is like a pump that returns the water to the uphill end of the circuit.

[14]Charge will accumulate more densely in the point, being attracted to the opposite charge across the gap. The charge density in turn affects the gradient of the electric potential across the gap, which is what causes the ionization. Therefore, approaching the doorknob with a flat hand can prevent the formation of an arc, and charge will simply flow (unnoticeably) after the contact has been made. This is also why lightning arresters work: a particularly pointed object like a metal rod will "attract" an electric arc toward its high charge density. By the same token, lightning tends to strike tall trees and transmission towers.

[15]The same is *not* true of electrical equipment that has been specifically designed to hold a very large amount of static charge!

[16]Unrelated to the EMF that stands for "Electromagnetic Fields."

When the wires are connected to form a complete loop, they make a *closed circuit*. If the wire were cut, this would create an *open circuit*, and the current would cease to flow. In practice, circuits are opened and closed by means of switches that make and break electrical contacts.

1.3.3 Voltage Drop

In describing circuits, it is often desirable to specify the voltage at particular points along the way. The difference in voltage between two points in a circuit is referred to as the *voltage drop* across the wire or other component in between. As in Ohm's law, $V = IR$, this voltage drop is proportional to the current flowing through the component, multiplied by its resistance.

As in the analogy of water pipes running downhill, the voltage drops continuously throughout a circuit, from one terminal of the emf to the other. However, just like the slope of the pipes may change, the voltage does not necessarily drop at a steady rate. Rather, depending on the resistance of a given circuit component, the voltage drop across it will be more or less: a component with high resistance will sustain a greater voltage drop, whereas a component with low resistance such as a conducting wire will have a smaller voltage drop across it, perhaps so small as to be negligible in a given context. For small circuits, it is often reasonable to assume that the wire's resistance is zero, and that therefore the voltage is the same all the way along the wire. In power systems, however, where transmission and distribution lines cover long distances, the voltage drop across them is significant and indeed accounts for some important aspects of how these systems function.

Importantly (and in contrast to the water analogy), the magnitude of the current also determines the voltage drop (along with resistance). For example, at times of high electric demand and thus high current flow, the voltage drop along transmission and distribution lines is greater; that is, the voltage drops more rapidly with distance. If this condition cannot be compensated for by other adjustments in the system (see Section 6.7), customers experience lower voltage levels associated with dimmer lights and impaired equipment performance, known as "brownouts." Similarly, if a piece of heavy power equipment is connected through a long extension cord with too high a resistance, the voltage drop along this cord can result in damage to the motor from excessively low voltage at the far end.

1.3.4 Electric Shock

Any situation where a high voltage is sustained by an electromotive force (or a very large accumulation of charge) constitutes a shock hazard. Our bodies are not noticeably affected by being "charged up," or raised to a potential above ground, just as birds can sit on a single power line. Rather, harm is done when a current flows *through* our body. A current as small as a few milliamperes across the human heart can be lethal.[17] For current to flow through an object, there must be a

[17] A saying goes, "It's the volts that jolts, and the mils that kills."

voltage drop across it. In other words, our body must be simultaneously in contact with two sources of different potential—for example, a power line and the ground.

Though it is the current that causes biological damage, Ohm's law indicates that shock hazard is roughly proportional to the voltage encountered. However, the resistance is also important. On an electrical path through the human body, the greatest resistance is on the surface of the skin and clothing, while our interior conducts very well. Thus, the severity of a shock received from a particular voltage can vary, depending on how sweaty one's palms are, or what type of shoes one is wearing.

The physical principles of electric current can be applied to suggest a number of practical precautions for reducing electric-shock hazards. For example, when touching an object at a single high voltage, we are safe as long as we are insulated from the ground. A wooden ladder might serve this purpose at home, while utility linemen often work on "hot" equipment out of raised plastic "buckets." Linemen can also insulate themselves from the high-voltage source by wearing special rubber gloves, which are commonly used for work on up to 12 kilovolts. The important thing is to know the capability of the insulator in relation to the voltage encountered.

A different safety measure often used by electricians when touching a questionable component (such as a wire that might be energized) is to make contact with ground potential with the same hand, for example by touching the little finger to the wall. In this way, a path of low resistance is created through the hand, which will greatly reduce the current flowing through the rest of the body and especially across the heart. Though the hand might be injured (improbable at household voltage), such a shock is far less likely to be lethal.

Around high-voltage equipment, in order to avoid the possibility of touching two objects at different potentials with a current pathway across the heart, a common practice is to "keep one hand in your pockets." Near very high potentials, where the concern is not just about touching equipment, but even drawing an arc across the air, the advice is to "keep both hands in your pockets" so as to avoid creating a point with high charge density to attract an arc.

Finally, another factor to consider is the muscular contraction that often occurs in response to an electric shock. Thus, a potentially energized wire is better touched with the back of the hand, so as to prevent involuntary closing of the hand around it.

If a person is in contact with an energized source, similar precautions should be exercised in removing them, lest there be additional casualties. If available, a device like a wooden stick would be ideal; in the worst case, kicking is preferable to grabbing.

1.4 RESISTIVE HEATING

Whenever an electric current flows through a material that has some resistance (i.e., anything but a superconductor), it creates heat. This *resistive heating* is the result of "friction," as created by microscopic phenomena such as retarding forces and collisions involving the charge carriers (usually electrons); in formal terminology, the heat corresponds to the work done by the charge carriers in order to travel to a lower potential. This heat generation may be intended by design, as in any heating appliance (for example, a toaster, an electric space heater, or an electric blanket). Such an appliance essentially consists of a conductor whose resistance is chosen so as to produce the desired amount of resistive heating. In other cases, resistive heating may be undesirable. Power lines are a classic example. For one, their purpose is to transmit energy, not to dissipate it; the energy converted to heat along the way is, in effect, lost (thus the term *resistive losses*). Furthermore, resistive heating of transmission and distribution lines is undesirable, since it causes thermal expansion of the conductors, making them sag. In extreme cases such as fault conditions, resistive heating can literally melt the wires.

1.4.1 Calculating Resistive Heating

There are two simple formulas for calculating the amount of heat dissipated in a resistor (i.e., any object with some resistance). This heat is measured in terms of *power*, which corresponds to energy per unit time. Thus, we are calculating a *rate* at which energy is being converted into heat inside a conductor. The first formula is

$$P = IV$$

where P is the power, I is the current through the resistor, and V is the voltage drop across the resistor.

Power is measured in units of *watts* (W), which correspond to amperes × volts. Thus, a current of one ampere flowing through a resistor across a voltage drop of one volt produces one watt of heat. Units of watts can also be expressed as joules per second. To conceptualize the magnitude of a watt, it helps to consider the heat created by a 100-watt light bulb, or a 1000-watt space heater.

The relationship $P = IV$ makes sense if we recall that voltage is a measure of energy per unit charge, while the current is the flow rate of charge. The product of current and voltage therefore tells us how many electrons are "passing through," multiplied by the amount of energy each electron loses in the form of heat as it goes, giving an overall rate of heat production. We can write this as

$$\frac{Charge}{Time} \cdot \frac{Energy}{Charge} = \frac{Energy}{Time}$$

and see that, with the charge canceling out, units of current multiplied by units of voltage indeed give us units of power.

The second formula for calculating resistive heating is

$$P = I^2 R$$

where P is the power, I is the current, and R is the resistance. This equation could be derived from the first one by substituting $I \cdot R$ for V (according to Ohm's law). As we discuss in Section 3.1, this second formula is more frequently used in practice to calculate resistive heating, whereas the first formula has other, more general applications.

As we might infer from the equation, the units of watts also correspond to amperes$^2 \cdot$ ohms ($A^2 \cdot \Omega$). Thus, a current of one ampere flowing through a wire with one ohm resistance would heat this wire at a rate of one watt. Because the current is squared in the equation, two amperes through the same wire would heat it at a rate of 4 watts, and so on.

Example

A toaster oven draws a current of 6 A at a voltage of 120 V. It dissipates 720 W in the form of heat. We can see this in two ways: First, using $P = IV$, $120\ \text{V} \cdot 6\ \text{A} = 720\ \text{W}$. Alternatively, we could use the resistance, which is $20\ \Omega$ ($20\ \Omega \cdot 6\ \text{A} = 120\ \text{V}$), and write $P = I^2R$: $(6\ \text{A})^2 \cdot 20\ \Omega = 720\ \text{W}$.

It is important to distinguish carefully how power depends on resistance, current, and voltage, since these are all interdependent. Obviously, the power dissipated will increase with increasing voltage and with increasing current. From the formula $P = I^2R$, we might also expect power to increase with increasing resistance, assuming that the current remains constant. However, it may be incorrect to assume that we can vary resistance without varying the current.

Specifically, in many situations it is the voltage that remains (approximately) constant. For example, the voltage at a customer's wall outlet ideally remains at 120 V, regardless of how much power is consumed.[18] The resistance is determined by the physical properties of the appliance: its intrinsic design, and, if applicable, a power setting (such as "high" or "low"). Given the standard voltage, then, the resistance determines the amount of current "drawn" by the appliance according to Ohm's law: higher resistance means lower current, and vice versa. In fact, resistance and current are inversely proportional in this case: if one doubles, the other is halved.

What, then, is the effect of resistance on power consumption? The key here is that resistive heating depends on the *square* of the current, meaning that the power is more sensitive to changes in current than resistance. Therefore, at constant voltage, the effect of a change in current outweighs the effect of the corresponding change in resistance. For example, decreasing the resistance (which, in and of itself, would tend to decrease resistive heating) causes the current to increase, which increases resistive heating by a greater factor. Thus, at constant voltage, the net effect of decreasing resistance is to increase power consumption. An appliance that draws more power has a lower internal resistance.

For an intuitive example, consider the extreme case of a *short circuit*, caused by an effectively zero resistance (usually unintentional). Suppose a thick metal bar were

[18]This is generally true because (a) changes in power consumption from an individual appliance are small compared to the total power supplied to the area by the utility, and (b) the utility takes active steps to regulate the voltage (see Section 6.6). Dramatic changes in demand do cause changes in voltage, but for the present discussion, it is more instructive to ignore these phenomena and treat voltage as a fixed quantity.

placed across the terminals of a car battery. A very large current would flow, the metal would become very hot, and the battery would be drawn down very rapidly. If a similar experiment were performed on a wall outlet by sticking, say, a fork into it, the high current would hopefully be interrupted by the circuit breaker before either the fork or the wires melted (DO NOT actually try this!). The other extreme case is simply an open circuit, where the two terminals are separate and the resistance of the air between them is infinite: here the current and the power consumption are obviously zero.

Example

Consider two incandescent light bulbs, with resistances of 240 Ω and 480 Ω. How much power do they each draw when connected to a 120 V outlet?

First we must compute the current through each bulb, using Ohm's law: Substituting $V = 120$ V and $R_1 = 240\ \Omega$ into $V = IR$, we obtain $I_1 = 0.5$ A. For $R_2 = 480\ \Omega$, we get $I_2 = 0.25$ A.

Now we can use these values for I and R in the power formula, $P = I^2R$, which yields $P_1 = (0.5\ \text{A})^2 \cdot 240\ \Omega = 60$ W and $P_2 = (0.25\ \text{A})^2 \cdot 480\ \Omega = 30$ W.

We see that at constant voltage, the bulb with twice the resistance draws half the power.

There are other situations, however, where the current rather than the voltage is constant. Transmission and distribution lines are an important case. Here, the reasoning suggested earlier does in fact apply, and resistive heating is directly proportional to resistance. The important difference between power lines and appliances is that for power lines, the current is unaffected by the resistance of the line itself, being determined instead by the load or power consumption at the end of the line (this is because the resistance of the line itself is very small and insignificant compared to that of the appliances at the end, so that any reasonable change in the resistance of the line will have a negligible effect on the overall resistance, and thus the current flowing through it). However, the voltage drop along the line (i.e., the difference in voltage between its endpoints, not to be confused with the line voltage relative to ground) is unconstrained and varies depending on current and the line's resistance. Thus, Ohm's law still holds, but it is now I that is fixed and V and R that vary. Applying the formula $P = I^2R$ for resistive heating with the current held constant, we see that doubling the resistance of the power line will double resistive losses. Since in practice it is desirable to minimize resistive losses on power transmission and distribution lines, these conductors are chosen with the minimal resistance that is practically and economically feasible.

1.4.2 Transmission Voltage and Resistive Losses

Resistive losses are the reason why increasingly high voltage levels are chosen for power transmission lines. Recalling the relationship $P = IV$, the amount of power transmitted by a line is given by the product of the current flowing through it and its voltage level (as measured either with respect to ground or between two lines or phases of one circuit). Given that a certain quantity of power is demanded, there is a choice as to what combination of I and V will constitute this power. A higher voltage level implies that in order to transmit the same amount of power, less current needs to flow. Since resistive heating is related to the square of the current, it is highly beneficial from the standpoint of line losses to reduce the current by increasing the voltage.

Before power transformers were available, transmission voltages were limited to levels that were considered safe for customers. Thus, high currents were required, causing so much resistive heating that it posed a significant constraint to the expansion of power transmission. With increasing power carried at a given voltage, an increasing fraction of the total power is lost on the lines, making transmission uneconomical at some point. The increase in losses can be counteracted by reducing the resistance of the conductors, but only at the expense of making them thicker and heavier. A century ago, Thomas Edison found the practical limit for transmitting electricity at the level of a few hundred volts to be only a few miles.

With the help of transformers that allow essentially arbitrary voltage conversion (see Section 6.3), transmission voltage levels have grown steadily in conjunction with the geographic expansion of electric power systems, up to about 1000 kilovolts (kV), and with the most common voltages around 100–500 kV. The main factor offsetting the economic benefits of very high voltage is the increased cost and engineering challenge of safe and effective insulation.

1.5 ELECTRIC AND MAGNETIC FIELDS

1.5.1 The Field as a Concept

The notion of a *field* is an abstraction initially developed in physics to explain how tangible objects exert forces on each other at a distance, by invisible means. Articulating and quantifying a "field" particularly helps to analyze situations where an object experiences forces of various strengths and directions, depending on its location. Rather than referring to other objects associated with "causing" such forces, it is usually more convenient to just map their hypothetical effects across space. Such a map is then considered to describe properties of the space, even in the absence of an actual object placed within it to experience the results, and this map represents the field.

For example, consider gravity. We know that our body is experiencing a force downward because of the gravitational attraction between it and the Earth. This gravitational force depends on the respective masses of our bodies and the Earth, but it also depends on our location: astronauts traveling into space feel less and less of a pull toward the Earth as they get farther away. Indeed, though the effect is small, we are even slightly "lighter" on a tall mountain or in an airplane at high altitude. If we were interested in extremely accurate measurements of gravity (for example, to calculate the exact flight path of a ballistic missile), we could construct a map of a "gravitational field" encompassing the entire atmosphere, which would indicate the strength of gravity at any point. This field is caused by the Earth, but does not explicitly refer to the Earth as a mass; rather, it represents in abstract terms the effect of the Earth's presence. The field also does not refer to any object (such as an astronaut) that it may influence, though such an object's mass would need to be taken into account in order to calculate the actual force on it. Thus, the gravitational field is a way of mapping the influence of the Earth's gravity throughout a region of space.

An alternative interpretation is to consider the field as a physical entity in its own right, even though it has no substance of its own. Here we would call gravity a property of *the space itself*, rather than a map telling us about objects such as the Earth *in* space. Indeed, the field itself can be considered a "thing" rather than a map, because it represents potential energy distributed over space. We know of the presence of this potential energy because it does physical *work* on objects: for example, a massive object within the field is accelerated, and in that moment, the energy becomes observable. With this in mind, we can understand the field as the answer to the question, Where does the potential energy reside while we are not observing it?

This notion of the field as a physical entity is a fairly recent one. Whereas classical physics relied on the notion of action-at-a-distance, in which only tangible objects figured as "actors," the study of very large and very small things in the 20th century has forced us to give up referring to entities that we can touch or readily visualize when talking about how the world works. Instead, modern physics has cultivated more ambiguity and caution in declaring the "reality" of physical phenomena, recognizing that what is accessible to our human perception is perhaps not a definitive standard for what "exists." Even what once seemed like the most absolute, immutable entities—mass, distance, and time—were proved ultimately changeable and intractable to our intuition by relativity theory and quantum mechanics.

Based on these insights, we might conclude that any quantities we choose to define and measure are in some sense arbitrary patterns superimposed on the vast web of energy and movement that constitutes reality, for the purpose of helping us apprehend this reality with our thoughts. In this sense, we are no more justified in considering a planet a "thing that really exists" than we are a gravitational field. What we really care about as scientists, though, is how useful such a conceptual pattern might be for describing the world in concise terms and making predictions about how things will behave. By this standard, the notion of a "field" does wonders. Physicists and engineers are therefore accustomed to regarding fields, however devoid of substance, as real, manipulable, and legitimate physical entities just like tangible objects. In any case, the reader should rest assured that it is quite all right to simply accept the "field" as a strange instrument of analysis that grows more palatable with familiarity.

1.5.2 Electric Fields

In Section 1.1, we characterized the electric potential as a property of the location at which a charge might find itself. A map of the electric potential would indicate how much potential energy would be possessed by a charge located at any given point. The *electric field* is a similar map, but rather of the electric *force* (such as attraction or repulsion) that would be experienced by that charge at any location. This force is the result of potential differences between locations: the more dramatically the potential varies from one point to the next, the greater the force would be on an electric charge in between these points. In formal terms, the electric field represents the *potential gradient*.

Consider the electric field created by a single positive charge, just sitting in space. Another positive charge in its vicinity would experience a repulsive force. This repulsive force would increase as the two charges were positioned closer together, or decrease as they moved father apart; specifically, the electric force drops off at a rate proportional to the square of the distance. This situation can be represented graphically by drawing straight arrows radially outward from the first charge, as in Figure 1.1a. Such arrows are referred to as *field lines*. Their direction indicates the direction that a "test charge," such as the hypothetical second charge that was introduced, would be pushed or pulled (in this case, straight away). The strength

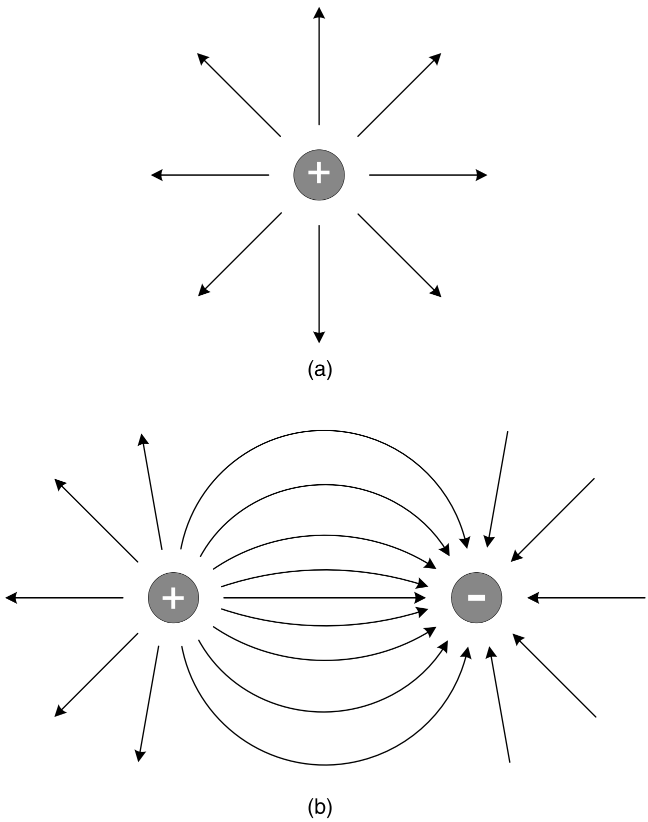

Figure 1.1 Electric field of (a) a single charge and (b) two opposite charges.

of the force is indicated by the proximity of field lines: the force is stronger where the lines are closer together.

This field also indicates what would happen to a negative charge: At any point, it would experience a force of equal strength (assuming equal magnitude of charge), but opposite direction as the positive test charge, since it would be attracted rather than repelled. Thus, a negative test charge would also move along the field lines, only backwards. By convention, the direction of the electric field lines is drawn so as to represent the movement of a *positive* test charge.

For a slightly more complex situation, consider the electric field created by a positive and a negative charge, sitting at a fixed distance from each other. We can map the field conceptually by asking, for any location, "What force would be acting on a (positive) test charge if it were placed here?" Each time, the net force on the test charge would be a combination of one attractive force and one repulsive force, in different directions and at different strengths depending on the distance from the respective fixed charges. Graphically, we can construct an image of the field by drawing an arrow in the direction that the charge would be pulled. The arrows for points along the charge's hypothetical path then combine into continuous field lines. Again, these field lines will be spaced more closely where the force is stronger. This exercise generates the picture in Figure 1.1b.

1.5.3 Magnetic Fields

The pattern of the electric field in Figure 1.1 may be reminiscent to some readers of the pattern that many of us produced once upon a time in science class by sprinkling iron filings on a sheet of paper over a bar magnet. The two phenomena, electric and magnetic forces, are indeed closely linked manifestations of a common underlying physics.

As we know from direct tactile experience, magnets exert force on each other: opposite poles attract, and like poles repel. This is somewhat analogous to the fact that opposite electric charges attract and like charges repel. But, unlike a positive or negative electric charge, a magnetic pole cannot travel individually. There is no such thing as an individual north or south pole (a "monopole" in scientific terms, which has never been found). Every magnet has a north and a south pole. Thus, unlike electric field lines that indicate the direction of movement of an individual test charge, magnetic field lines indicate the *orientation* of a test magnet. The iron filings in the familiar experiment—which become little test magnets since they are magnetized in the presence of the bar magnet—do not move toward one pole or the other, but rotate and align themselves with the direction of the field lines.

It is important to emphasize that, despite the similar shape of field lines, magnetic poles are not analogous to single electric charges sitting in space. Rather than thinking of magnetism as existing in the form of "stuff" like electric charge (which could conceivably be decomposed into its "north" and "south" constituents), it is more appropriate to think of magnetism as an expression of *directionality*, where north is meaningless without south. If you cut a magnet in half, you get two smaller magnets that still each have a north and a south pole.

If we pursued such a division of magnets again and again, down to the level of the smallest particles, we would find that even individual electrons or protons appear as tiny magnets. In ordinary materials, the orientation of all these microscopic magnets varies randomly throughout space, and they therefore do not produce observable magnetic properties at the macroscopic level. It is only in *magnetized* materials that the direction of these myriad tiny magnets becomes aligned, allowing their magnetic fields to combine to become externally noticeable. This alignment stems from the force magnets exert on each other, and their resulting tendency to position themselves with their north poles all pointing in the same direction. Some substances like *magnetite* occur naturally with a permanent alignment, making the familiar magnets that adhere to refrigerators and other things. Other materials like iron and steel can be temporarily magnetized in the presence of a sufficiently strong external magnetic field (this is what happens to the refrigerator door underneath the magnet), with the particles returning to their disordered state after the external field is withdrawn.

The magnetic property of microscopic particles is due to their electric charge and their intrinsic motion, which brings us to the fundamental connection between electricity and magnetism. Indeed, we can think of magnetism as nothing but a manifestation of directionality associated with electric charge in motion, whereby moving charges always exert a specific directional force on other moving charges. At the level of individual electrons, their motion consists of both an orbital movement around the atom's nucleus and an intrinsic *spin*, which we can visualize as if the particle were spinning like a top.[19] Both of these rotational motions combine to form what is referred to as a *magnetic moment*. Similarly, the protons inside atomic nuclei possess a magnetic moment due to their intrinsic spin.[20]

Knowing this, it would stand to reason that a large amount of moving charge such as a measurable electric current should produce a magnetic field as well. This phenomenon was in fact discovered in 1820, when Hans Christian Oersted observed that a compass needle was deflected by an electric current through a nearby wire. The magnetic field produced by an electric current points at a right angle to the flow of charge, in a direction specified by the "right-hand rule" illustrated in Figure 1.2. If the thumb of one's right hand is pointing in the direction of the current, then the curled fingers of the same hand indicate the direction of the magnetic field. Thus, the magnetic field lines surround the wire in a circular manner.

In order to make practical use of this phenomenon, we can alter the shape of the current-carrying wire by winding it into a coil, which brings many turns of wire closely together so that their magnetic fields will add to form a "straight" field in the center of the coil that is comparable to that of a bar magnet (an illustration of such a coil and its magnetic field is shown in Figure 3.4). This arrangement can be thought of as "concentrating" the magnetic field in space.

[19]Such a mechanistic representation is not quite accurate in terms of quantum mechanics, but it is nonetheless useful for constructing some intuitive picture.

[20]This effect is quite subtle and not important in our context, but is exploited in such technologies as Magnetic Resonance Imaging (MRI) for medical diagnostic purposes, which discriminates among tissues of different water content by way of the magnetic properties of the hydrogen nucleus.

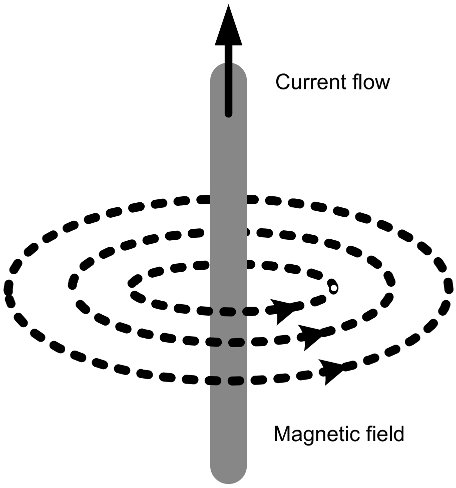

Figure 1.2 Magnetic field around a current-carrying wire.

Magnetic effects are essential for the generation and conversion of electric power. In order to successfully navigate the literature on these applications, it is important to be aware of a distinction between two types of quantities: one is called the *magnetic field* and the other *magnetic flux*. Despite the earlier caution, it is at times helpful (and indeed consistent with the Latin translation) to think of the flux as the directional "flow" of something, however immaterial, created in turn by the flow of electric current. Conceptually as well as mathematically, the flux is a very convenient quantity for analyzing electrical machines, while the magnetic field is particularly useful for describing the basic principles of electromagnetic induction in simplified settings.

Conventionally, the magnetic field is denoted by the symbol B and measured in units of *tesla* (T) or *gauss* (G). One tesla, which equals 10,000 or 10^4 gauss, corresponds to one newton (N) (a measure of force) per ampere (current) per meter: $1\ \text{T} = 1\ \text{N/A-m}$. Magnetic flux is denoted by ϕ (the Greek phi) and is measured in units of weber (Wb). One tesla equals one weber per square meter.

From this relationship between the units of flux and field, we can see that the magnetic field corresponds to the density or concentration in space of the magnetic flux. The magnetic field represents magnetic *flux per unit area*. Stated in reverse, magnetic flux represents a measure of the magnetic field multiplied by the area that it intersects.

Unless "concentrated" by a coil, the magnetic field associated with typical currents is not very strong. For example, a current of 1 ampere produces a magnetic field of 2×10^{-7} T or 0.002 G (2 milligauss) at a distance of 1 meter. By comparison, the strength of the Earth's magnetic field is on the order of half a gauss.[21]

[21]The exact value of the Earth's magnetic field depends on geographic location, and it is less if only the horizontal component (to which a compass needle responds) is measured.

1.5.4 Electromagnetic Induction

While electric current creates a magnetic field, the reverse effect also exists: magnetic fields, in turn, can influence electric charges and cause electric currents to flow. However, there is an important twist: the magnetic field must be *changing* in order to have any effect. A static magnetic field, such as a bar magnet, will not cause any motion of nearby charge. Yet if there is any *relative* motion between the charge and the magnetic field—for example, because either the magnet or the wire is being moved, or because the strength of the magnet itself is changing—then a force will be exerted on the charge, causing it to move. This force is called an *electromotive force* (*emf*) which, just like an ordinary electric field, is distinguished by its property of accelerating electric charges.

The most elementary case of the electromotive force involves a single charged particle traveling through a magnetic field, at a right angle to the field lines (the direction along which iron filings would line up). This charge experiences a force again at right angles to both the field and its velocity, the direction of which (up or down) depends on the sign of the charge (positive or negative) and can be specified in terms of another right-hand rule, as illustrated in Figure 1.3.

This effect can be expressed concisely in mathematical terms of a *cross product* of vector quantities (i.e., quantities with a directionality in space, represented in boldface), in what is known as the Lorentz equation,

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$$

where **F** denotes the force, q the particle's charge, **v** its velocity, and **B** the magnetic field. In the case where the angle between **v** and **B** is 90° (i.e., the charge travels at right angles to the direction of the field) the magnitude or numerical result for **F** is simply the arithmetic product of the three quantities. This is the maximum force

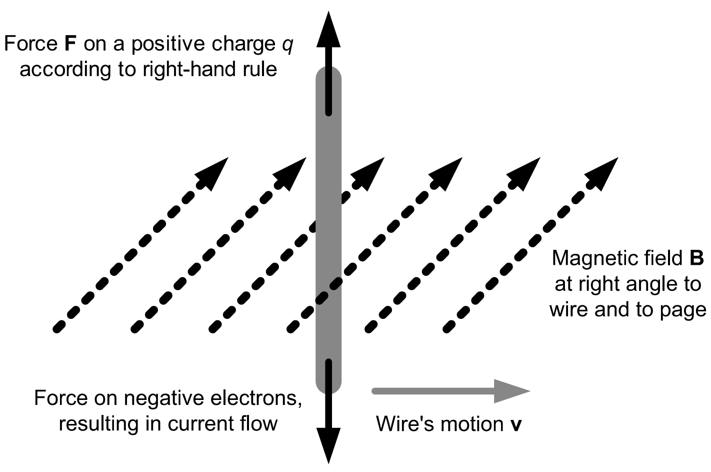

Figure 1.3 Right-hand rule for the force on a charge.

possible: as the term cross product suggests, the charge has to move *across* the field in order to experience the effect. The more **v** and **B** are at right angles to each other, the greater the force; the more closely aligned **v** and **B** are, the smaller the force. If **v** and **B** are parallel—that is, the charge is traveling along the magnetic field lines rather than across them—the force on the charge is zero.

Figure 1.3 illustrates a typical application of this relationship. The charges q reside inside a wire, being moved as a whole so that each of the microscopic charges inside has a velocity **v** in the direction of the wire's motion. If we align our right hand with that direction **v** and then curl our fingers in the direction of the magnetic field **B** (shown in the illustration as pointing straight back into the page), our thumb will point in the direction of the force **F** on a positive test charge. Because in practice the positive charges in a metal cannot move but the negatively charged electrons can, we observe a flow of electrons in the negative or opposite direction of **F**. Because only the *relative* motion between the charge and the magnetic field matters, the same effect results if the charge is stationary in space and the magnetic field is moved (e.g., by physically moving a bar magnet), or even if both the magnet and the wire are stationary but the magnetic field is somehow made to become stronger or weaker over time. As we will see in Chapter 4, a combination of these effects—movement through space of wires and magnets, as well as changing magnetic field strength—is employed in the production of electric power by generators.

The phenomenon of *electromagnetic induction* occurs when this electromagnetic force acts on the electrons inside a wire, accelerating them in one direction along the wire and thus causing a current to flow. The current resulting from such a changing magnetic field is referred to as an *induced current*. This is the fundamental process by which electricity is generated, which will be applied over and over within the many elaborate geometric arrangements of wires and magnetic fields inside actual generators.

1.5.5 Electromagnetic Fields and Health Effects

A current flowing through a wire, alternating at 60 cycles per second (60 Hz), produces around it a magnetic field that changes direction at the same frequency. Thus, whenever in the vicinity of electric equipment carrying any currents, we are exposed to magnetic fields. Such fields are sometimes referred to as *EMF*, for *electromagnetic fields*, or more precisely as *ELF*, for *extremely low-frequency fields*, since 60 Hz is extremely low compared to other electromagnetic radiation such as radio waves (which is in the megahertz, or million hertz range).

There is some concern in the scientific community that even fields produced by household appliances or electric transmission and distribution lines may present human health hazards. While such fields may be small in magnitude compared to the Earth's magnetic field, the fact that they are oscillating at a particular frequency may have important biological implications that are as yet poorly understood.

Research on the health effects of EMFs or ELFs continues. Some results to date seem to indicate a small but statistically significant correlation of exposure to ELFs from electric power with certain forms of cancer, particularly childhood leukemia, while other studies have found no effects.[22] In any case, the health effects of ELFs on adults appear to be either sufficiently mild or sufficiently rare that no obvious disease clusters have been noted among workers who are routinely exposed—and have been over decades—to vastly stronger fields than are commonly experienced by the general population.

From a purely physical standpoint, the following observations are relevant: First, the intensity of the magnetic field associated with a current in a wire is directly proportional to the current; second, the intensity of this field decreases at a rate proportional to the inverse square of the distance from the wire, so that doubling the distance reduces the field by a factor of about 4. The effect of distance thus tends to outweigh that of current magnitude, especially at close range where a doubling may equate to mere inches. It stands to reason, therefore, that sleeping with an electric blanket or even an electric alarm clock on the bedside table would typically lead to much higher exposure than living near high-voltage transmission lines. Measured ELF data are published by many sources.

1.5.6 Electromagnetic Radiation

Although not vital in the context of electric power, another manifestation of electromagnetic interactions deserves at least brief discussion: namely, electromagnetic waves or *radiation*, including what we experience as *light*. Visible light in fact represents a small portion of an entire spectrum of electromagnetic radiation, which is differentiated by *frequency* or *wavelength*. The nonvisible (to us) regions of this spectrum include infrared and ultraviolet radiation, microwaves, radio waves and others used in telecommunications (such as cellular phones), X rays, as well as gamma rays from radioactive decay. Physically, all these types of radiation are of the same basic nature.

The concept of a wave is familiar to most of us as the periodic movement of some material or medium: the surf on a Hawaiian beach, a vibrating guitar string, or the coordinated motion of sports fans in the bleachers. What is actually "waving" when electromagnetic radiation travels through space is much less tangible and challenging to the imagination; we can describe it only as a pulse of rapidly increasing and decreasing electric and magnetic fields, themselves completely insubstantial and yet measurably affecting their environment. These electric and magnetic fields are at right angles to each other, and at right angles again to the direction of propagation of the wave, as illustrated in Figure 1.4.

The *frequency* of the electromagnetic wave refers to the rate at which either the electric or magnetic field at any one point oscillates (i.e., changes direction back and

[22]Technical information and summaries of research have been published by the World Health Organization, http://www.who.int/mediacentre/factsheets/ (accessed November 2004), and by the National Institutes of Health, http://www.niehs.nih.gov/emfrapid/home.htm (accessed November 2004). See, for example, C.J. Portier and M.S. Wolfe (eds.), *Assessment of Health Effects from Exposure to Power-Line Frequency Electric and Magnetic Fields*, NIEHS Working Group Report (Research Triangle Park, NC: National Institute of Environmental Health Sciences of the National Institutes of Health, 1998).

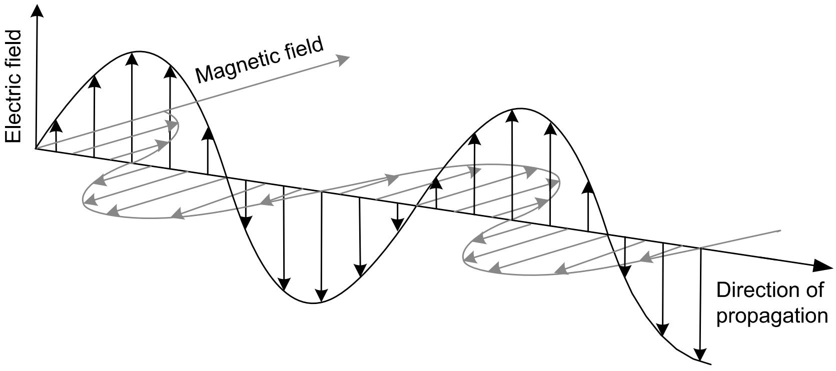

Figure 1.4 An electromagnetic wave.

forth). Frequency is measured in cycles per second, or *hertz* (Hz). The *wavelength* represents the distance in space from one wave crest to the next (analogously, the *period* of a wave represents the separation in time from one wave crest to the next). Depending on the range of the spectrum in question, wavelength may be measured in meters or any small fraction thereof. For example, the wavelength of a certain color of visible light might be quoted in microns ($1\ \mu m = 10^{-6}$ m), nanometers ($1\ nm = 10^{-9}$ m), or angstroms ($1\ \text{Å} = 10^{-10}$ m).

Wavelength is inversely proportional to the frequency; higher frequency implies shorter wavelength, and vice versa. This is because wavelength and frequency multiplied together yield the speed of propagation of the wave, which is fixed: the speed of light. Aside from the detail that the speed of an electromagnetic wave actually varies slightly depending on the medium through which it is traveling, the constancy of the speed of light is famously important.[23]

This constancy can be understood as a manifestation of the principle of the *conservation of energy*: It is at this speed and only at this speed of propagation, or rate of change of electric and magnetic fields, that they keep inducing each other at the same magnitude. Were the wave to propagate more slowly, the fields would decay (implying energy that mysteriously vanishes); were it to propagate any faster, the fields would continually increase (implying a limitless creation of energy). As we know from the first law of thermodynamics, energy can be neither created nor destroyed in any physical process. From this basic principle it is possible to derive the constant speed of light.

Electromagnetic radiation interacts with matter through charges—specifically, electrons—that are accelerated and moved by the field. Let us consider first the example of radio waves that are broadcast and received through conducting metal

[23]In the fine print of the physics textbook we learn that 3×10^8 meters per second, or 186,000 miles per hour, is the speed of light only *in a vacuum*. Light travels somewhat slower through various materials, and these small differences in the speed of light—both as a function of the medium and of wavelength—give rise to familiar optical phenomena like refraction in a lens, prism, or glass of water.

antennas, and then the more general case of photon absorption and emission in all types of materials.

As the music plays at the radio station, its specially encoded electronic signal travels in the form of a rapidly changing electric current into the station's large antenna, moving the electrons inside the metal up and down.[24] These moving electrons produce a pulse of a changing electric field that is "felt" in the region of space surrounding the antenna. This oscillating electric field induces a magnetic field, which in turn induces an electric field, and so forth, with the fields propagating away in the form of a wave—an electromagnetic wave of a very specific time signature. The wave becomes weaker with increasing distance from the antenna in that it spreads out through space, though the "pulse" itself is preserved as long as the wave is detectable.

Another antenna at a distance can now "receive" the wave because the electrons inside it will be accelerated by the changing electric field, in exactly the same fashion as the electrons responsible for "sending" it. We can see that an antenna needs to be conducting so as to allow the electrons to move freely according to the changing field. When this induced motion of electrons is decoded by the radio receiver, the electric signal travels through a wire and finally moves the magnet of a loudspeaker back and forth, the specific signature of the electromagnetic wave is translated back into sound.

This large-scale motion of electrons, as in radio antennas, is a special case of their interaction with electromagnetic radiation. More generally, electrons stay within their atomic orbitals, but they nevertheless undergo certain transitions that allow them to "send out" or "receive" electromagnetic radiation. These transitions are not readily represented in terms of physical motion, but can only be described in the language of quantum mechanics. Physicists say that an electron changes its *energy level*, and that the difference between the energy level before and after the transition corresponds to the energy carried by a "packet" of electromagnetic radiation. Such a packet is called a *photon*.[25] As an electron moves to a state of lower energy, it emits a photon, and conversely, as it absorbs a photon, it rises to a state of higher energy.

Although the photon itself has no mass and can hardly be conceived of as an "object," it is nonetheless transporting energy through space. The amount of energy is directly proportional to the frequency of the radiation: the higher the frequency, the greater the energy. Thus, we can think of the "waving" electric and

[24]There are two standard types of encoding: amplitude modulation (AM) and frequency modulation (FM). In each case, the sound signal (which itself is an electrical pulse of changing voltage and current that mimics the corresponding sound wave) is superimposed on a *carrier wave* of a given frequency (the broadcast frequency of that particular station, which, at many kilohertz or megahertz, is several orders of magnitude higher than the frequency of the signal). In AM, the amplitude of the carrier wave is continually changed (modulated) according to the signal; in FM, the frequency is changed by a small percentage.

[25]It was one of the most stunning discoveries in early 20th-century physics that radiation occurs in such packets, or *quanta*, that only interact with a single electron at a time; the crucial experiment that demonstrated this (the *photoelectric effect*) is what actually won the 1921 Nobel Prize in Physics for Einstein.

magnetic fields in space as a form of potential energy, whose presence does not become apparent until it interacts with matter.

The configuration of electrons within a given material, having a certain atomic and molecular structure, determines what energy transitions are available to electrons. They will interact with radiation only to the extent that the available transitions match precisely the energy of the photon, corresponding to its frequency (wavelength). This explains why materials interact differently with radiation of different frequencies, absorbing some and transmitting or reflecting others. A glass window, for example, transmits visible light but not ultraviolet. And we find ourselves—at this very moment—in a space full of radio waves, oblivious to their presence because the waves pass right through our bodies: the energy of their individual photons is insufficient to cause a transition inside our electrons.

In the context of electric power system operation, electromagnetic radiation does not play much of an explicit role. This is because the conventional frequency of alternating current at 50 or 60 hertz is so low that the corresponding radiation propagates with extremely little energy and is in practice unobservable. Stationary and alternating electric and magnetic fields, however, are central to the workings of all electric machinery.