History of EM

Home | Up | Photography | Games | Opera

History of EM | 35mm Slides

The History of Electromagnetics

Electricity, magnetism, and light surround us inescapably. Of the four fundamental forces known to modern physics—*electromagnetism*, *gravity*, the *weak nuclear force*, and the *strong nuclear force*—we can manipulate only electromagnetism on a human scale. We cannot perceive the gravitational force between two objects that we can hold in our hands; only with objects of astronomical sizes are the gravitational forces large enough to be felt, and those we cannot change. The weak and strong nuclear forces operate at tiny distances measured in femtometers (1,000,000,000,000,000 femtometers = 1 meter) that are inaccessible to unaided human observation. All of the forces that we can control with our bodies, including the chemical reactions that send signals along our nervous systems and power our muscles, are fundamentally electromagnetic.

Because electromagnetic phenomena are directly accessible to human observation and manipulation, electricity and magnetism were known in ancient times, although not well understood. While a deep relationship between electricity and magnetism was long suspected, it was only in the nineteenth century that physicists developed a theoretical and mathematical basis that tied them together. The theory of electromagnetics was the preeminent success story of nineteenth-century physics. Not only did it specify the precise relationship between electricity and magnetism, it predicted the existence of electromagnetic waves and showed that light is simply one type of electromagnetic wave. This insight led directly to the development of radio and other forms of electromagnetic communication. For most applications in electrical engineering, we still use a refined version of the mathematical formulation of electromagnetic theory published by James Clerk Maxwell in 1873, known as *Maxwell's equations*.

Further developments of electromagnetic theory in the twentieth century took place in four general areas. First, early in the twentieth century, Albert Einstein realized that Maxwell's equations contained a startling implication: space and time are aspects of a single quantity that we now call *space-time*. Different observers, traveling at different velocities in different frames of reference, will not agree on the space and time intervals between two events. They will only agree on an aggregate quantity called the *space-time interval*. Similarly, the two observers in different frames of reference will not agree on the separate values of the electric and magnetic fields, but only on a combination of both called the *electromagnetic field tensor*. Electricity and magnetism are therefore not only deeply related, but are actually aspects of a single underlying phenomenon. (We will define the space-time interval and the electromagnetic field tensor in later chapters. For now, we can think of them simply as combinations of their respective individual parts.)

In the late twentieth century, physicists continued to advance the idea of unification of fundamental forces by first developing a unified theory of electromagnetics and the weak nuclear force. The unified result is called the *electoweak* force. Recent (but still incomplete) theories add the strong nuclear force as well. Gravity alone has eluded attempts at a *grand unification* of all forces.

The second area of modern research is the interaction of electromagnetic fields with matter. It is only through these interactions that electromagnetic fields make their presence tangible. For example, electromagnetic waves in the form of light cause chemical reactions in our eyes that allow us to see. At lower frequencies, magnetic forces in motors convert electromagnetic energy to mechanical energy. While nineteenth century physicists had a basic understanding of the interactions of electric and magnetic fields with materials at the macroscopic or bulk level, a detailed and experimentally verified model of the electromagnetic behavior of matter at the atomic level was not available until the twentieth century.

The third area of research in electromagnetic theory in the late nineteenth and twentieth centuries was the compilation of mathematical analysis techniques to solve Maxwell’s equations in a variety of configurations. Before the advent of high-speed digital computers, these were the only practical means of making quantitative predictions using Maxwell’s equations. Typically, an electromagnetic analyst would study a problem, reduce its complexity to a few essential features, and then express the solution as a known mathematical function that satisfies Maxwell’s equations for the simplified problem. These techniques are also called *closed-form* methods, because they lead to the solution’s expression as an exact mathematical function. For example, to calculate the electromagnetic waves scattering from an object, one could model the object as a sphere and then express the electromagnetic field as a mathematical function that satisfies Maxwell’s equations in all space as well as the correct boundary conditions for Maxwell’s equations at the surface of the sphere.

The fourth area of development in electromagnetics in the twentieth century was the advent of high-speed digital computers capable of solving Maxwell’s equations approximately. These approximate or *numerical* methods have the advantage of allowing for a broader range of problems to be solved than closed-form methods. Numerical methods reduce—but do not eliminate—the amount of abstraction needed to solve a problem. Going back to our example of electromagnetic waves scattering from an object, with numerical methods it is possible to model an object with an arbitrary shape. However, in order to reduce the difficulty of solving the problem with limited computing resources, we must identify and eliminate non-essential features. In addition, we usually must accept that numerical solutions will be less accurate than closed-form solutions in cases where both can be calculated.

Whether we solve electromagnetics problems with closed-form or numerical techniques, it is essential for us first to understand the *qualitative* (non-numerical) nature of the underlying physical electromagnetic phenomena. Armed with that understanding, we will be ready to apply the appropriate solution techniques to obtain *quantitative* (numerical) values for the electromagnetic quantities of interest for our particular problems.

Despite the many false starts, missteps, and blind alleys encountered by researchers in electromagnetics, history provides an excellent guide and foundation to the study of electromagnetic phenomena. The remainder of this chapter will discuss the history of electromagnetics in detail, concentrating on the researchers who advanced the subject and their key discoveries. For convenience, we divide the history of electromagnetics into three periods: **Ancient Era**, **Scientific Revolution**, and **Modern Era**.

The Ancient Era

The Ancient Era of scientific investigation began at the dawn of human civilization, reaching its zenith with the Greek philosophers. Most science of the ancient era was *qualitative*—describing the general nature of a phenomenon—rather than *quantitative*—measuring the phenomenon. For example, the fact that amber when rubbed attracts light objects was known to the scientists of the Ancient Era, but they did not measure the force of attraction and relate it to factors such as the types of objects attracted and the distances from the amber to the objects.

The lack of quantitative measurements in ancient science can be explained only partially by the limited availability of accurate measurement instruments. Balances were widely used by merchants in the ancient world to measure weights of large and small objects. Reasonably accurate length measurements were essential to city planning and architecture. Ancient astronomical instruments such as the astrolabe measured the locations of celestial objects to a small fraction of a degree. If the ancients could accurately measure forces (weights), lengths, and angles, why did quantitative measurements of electromagnetic forces not begin in the ancient world? The answer is that ancient scientists lacked the conceptual framework of the scientific method to motivate them to perform the measurements. We will trace the development of the scientific method in our discussion of the Scientific Revolution later in this chapter.

The earliest scientist known with certainty to have discussed electricity and magnetism was the Greek philosopher, Thales (625? –546? BC). No writings of his own survive. What we know of Thales comes principally from commentaries of Greek historian Herodotus (484? –425 BC) and Greek philosopher Aristotle (384–322 BC). Thales’ fame was founded on his prediction of the eclipse of the sun on May 28, 585 BC. The eclipse was used as a convenient excuse to end an inconvenient war between the Lydians and the Medes. He is also credited with introducing Egyptian geometric ideas to Greece.

Thales knew that the mineral *magnetite* (named after the district of *Magnesia* in Asia Minor, where it was first discovered in quantity), also known as lodestone, exerts an attractive force on nearby iron objects. In *On the Soul*, Aristotle states that

> *Thales, too, to judge from what is recorded about him, seems to have held soul to be a motive force, since he said that the magnet has a soul in it because it moves the iron.*

Thales’ ascribing a soul to a magnet to explain its ability to attract iron seems quaint today. However, we should remember that the concept of natural phenomena regulated by natural laws rather than the arbitrary whims of the gods was only beginning to take hold in Greece at the time of Thales. The idea of abstract natural laws behind physical events found further expression in the teachings of philosophers who followed Thales, particularly Aristotle.

The science of magnetism, meaning the systematic study of magnetic properties, truly began with a remarkable letter written by Peter Peregrinus (born Pierre de Maricourt) from his army encampment during the siege of Lucera, Italy, on August 8, 1269. In his letter, Peregrinus described his experiments with a sphere of magnetite. He discovered that the sphere had two spots toward which an iron needle pointed and he coined the term *polus* (poles) for these spots. He studied the attraction and repulsion of two floating lodestones, and described how they tend to point north-south along their poles

Most of his letter dealt with schemes to create a perpetual motion machine with magnets. Again, from today's perspective this might seem humorous, although misguided attempts to construct perpetual motion machines from permanent magnets persist to this day. However, without having a scientific theory of magnetism, it was reasonable to investigate experimentally whether the magnetic force might be used to drive continuous motion. Peregrinus was well ahead of his time—it would take six centuries for the idea of continuous motion driven by magnetic force (and obeying the law of conservation of energy!) to find practical application in electric motors.

While the ability of amber to attract small objects was only a scientific curiosity in ancient times, the tendency of magnetite (or magnetized iron needles) to align north-south was a useful property, known both in the East and West prior to the Scientific Revolution. Being able to find north without recourse to astronomical observations has obvious benefits to travelers, especially mariners in foggy or cloudy weather. Because of this useful property, the craft of compass making became a fine art in the seafaring nations of Europe.

English instrument maker Robert Norman published his observations concerning compass needles in his 1581 book, *The New Attractive*. It was known by then that compass needles do not point to true north as determined by astronomical observations. Some ascribed the reason for this to improper methods of magnetizing the needle. Norman discovered an even more disturbing fact—the compass needle not only fails to point to true north, but also tilts with respect to the horizontal. Other compass makers knew of this effect because it requires a counterweight on the needle to keep it horizontal, but Norman was the first to make detailed observations and publish them. He determined through a series of ingenious experiments that there was no net force acting on the needle to pull it up or down, but only a force turning it to tilt at a particular angle, which he termed *The Line Respective*. He incorrectly ascribed the tilt effect to the original lodestone used to magnetize the needle, but admitted that he could not explain its cause.

William Gilbert (1544–1603) is a transitional figure from the Ancient Era of scientific research to the Scientific Revolution. He published in 1600 a book that marks the beginning of the transition: *Of Magnets, Magnetic Bodies, and the Great Magnet of the Earth*. In this book, Gilbert debunked many folk superstitions regarding the imagined magical properties of magnets, such as their supposed abilities to detect marital infidelity, cure headaches, and find gold. (Belief in the magical curative power of magnets persists today.) He used the observed horizontal tilt of compass needles and his own experiments with iron needles near a magnetite sphere to deduce that the earth itself must be a large magnet. He discovered that a magnetized piece of iron loses its magnetization when heated to incandescence, and that conversely an unmagnetized piece of wrought iron can be magnetized by striking it with a hammer. Gilbert applied the term *electric* force (from the Greek word for amber, *elektron*) to the force that attracts small objects to amber and other substances after rubbing. He distinguished that force from *magnetic* force (from the mineral *magnetite*), which is intrinsic to certain materials.

As is often the case in science, not all of the conclusions Gilbert drew from his experiments were correct. A firm believer in the Copernican theory of the solar system, he postulated that magnetism provided the force that kept the planets in their orbits and kept the earth spinning on its axis. Newton would later show that gravity was responsible for the orbits of the planets, and that once the earth started spinning on its axis (presumably at its creation), no additional force was required to keep it spinning.

Nevertheless, Gilbert's research marked the beginning of two powerful ideas that influenced the scientists who followed.

First, Gilbert established that scientific treatises should be based upon extensive observations and experiments, from which the conclusions would then be reached. In this belief, Gilbert followed in the tradition of English scientist and philosopher Roger Bacon (1214? –1294), who was declared a heretic in his own time. While it seems obvious today, observation and experiment stood in contrast to the established tradition in Gilbert's time of reliance on ancient authorities and logical arguments. For example, many educated people accepted the following explanation of the motions of the planets in the sky: Plato taught that the circle was the most perfect geometric form, and since the heavens must be perfect, therefore planets must travel in circles—if not in simple circles like the stars (observations of the planets disproved this), then in smaller circles superimposed on larger circles, which were called *epicycles*. The inability of epicycles to account for the motions of the planets helped to end this faulty mode of scientific investigation.

The second principle established by Gilbert's work was the counter-intuitive idea of action at a distance. It is our common human experience that we must touch things in order to exert force on them. The idea that the sun might exert force on a planet from a distance became easier to accept after Gilbert's studies of magnetism were published, although it turned out that magnetism was not the force governing planetary orbits. Ironically, physics has come full circle on this issue. Present theories emphasize the concept of force between two objects being mediated by the exchange of real or virtual particles rather than an abstract force at a distance.

The Scientific Revolution

The Scientific Revolution brought a change in electromagnetics research from qualitative descriptions to quantitative measurements. In order to effect this change, two developments were needed: (1) devices for creating and measuring electromagnetic quantities, and (2) mathematical methods to describe the measurements and allow predictive calculations for experiments not yet performed. We begin our discussion of the Scientific Revolution with the mathematical developments of Isaac Newton (1642–1727) and Gottfried Wilhelm Leibniz (1646–1717).

The development of a quantitative theory of electromagnetics required a new type of mathematics—*differential calculus*—capable of expressing with great precision and elegant simplicity the relationships yet to be discovered governing electric and magnetic phenomena. While differential calculus did not immediately change scientific investigation—indeed, Faraday was able to advance the science of electromagnetism without the benefit of advanced mathematics—ideas from calculus gradually found their way into science until differential and integral calculus became an essential component of electromagnetic theory by the time of Maxwell.

Newton and Leibniz share the honor of the discovery of differential calculus. Newton invented differential calculus (which he called the *fluxional method*) in 1666 in order to solve the problems of finding the tangents to curves and the areas under curves. Fearful of criticism, Newton did not publish his method until 1687. Leibniz independently discovered differential calculus (his name for the method) in 1675 and published his findings in 1684. Because Leibniz published first and introduced the method to Europe, his name for the method and his notations are generally used today, although Newton's notation for the derivative is sometimes used as well.

With his publication of *Philosophiae Naturalis Principia Mathematica* in 1687, Newton began the Scientific Revolution in earnest. This book set forth the fundamental principles of theoretical mechanics that were deemed immutable laws of the universe until the twentieth century. Prior to Newton's work, Johannes Kepler (1571–1630) had rejected epicycles and devised three laws of planetary motion based on astronomical observations. Kepler recognized that these laws implied the existence of a force exerted by the sun on the planets that decreased with distance, but he lacked the mathematical techniques to make this idea more precise. Newton showed that the force of gravity between two bodies obeys an inverse square dependence on distance:

$$F = \frac{G m_1 m_2}{r^2} \qquad (1.1)$$

where F is the force, G is a universal constant, m1 is the mass of one body, m2 is the mass of the other, and r is the distance between them. This relationship, coupled with Newton's laws of motion, explained Kepler's laws of planetary orbits.

Newton's expression of his laws as equations in differential calculus set the standard for Maxwell's equations two centuries later. Prior to Newton and Leibniz, physical relationships were usually given in geometric terms when they were described mathematically. For example, Kepler's second law of planetary motion states that equal areas are swept out in equal times as a planet orbits along the path of an ellipse. Newton's expression of the same law as an equation in differential calculus was less intuitive but facilitated the law's extension to the more general case of an arbitrary body in motion subjected to an arbitrary force. This gave physical theory a predictive power it did not have previously. Kepler's laws described motions that had already been observed. Newton's laws were capable not only of describing motions that had already been observed, but of predicting motions that had not yet been observed.

The next historical step in the development of a complete theory of electromagnetics was to create and measure electromagnetic quantities in the laboratory. The first practical method for producing large quantities of electric charge was invented in 1660 (before the publication of Newton's *Principia*) by Otto von Guericke (1602–1686). His generator consisted of a sulfur sphere mounted on a shaft connected to a hand crank. The operator spun the sulfur sphere with the crank while pressing a silk cloth to its surface. Brushes picked up the electric charge. With this portable device, an electric charge could easily be delivered to any object the experimenter desired.

Peter van Musschenbrock (1692–1761) improved this idea with the accidental discovery in 1745 of what we now call a capacitor or condenser—a device capable of storing electric charge for extended periods. The present form of his device consists of a glass jar lined with metal foil on the inside and outside, with a wire leading through the neck to the inside. This device is known as a *Leyden jar* after the University of Leyden, Netherlands, where van Musschenbrock did his research.

Figure 1.1—Leyden Jar

A presentation of the history of electromagnetics would be incomplete without a brief mention of Benjamin Franklin (1706–1790). Franklin's famous 1752 kite experiment demonstrated that lightning is a form of electricity. Franklin also invented the lightning rod. Most importantly for the development of electromagnetic theory, Franklin proposed an explanation for the apparent existence of two kinds of electricity, *positive* and *negative*. The two types of electric charges were first identified in 1733 by French scientist Charles François de Cisternay Du Fay. Du Fay believed that there were two different electrical fluids, positive and negative. Franklin formulated an alternate explanation in 1747—a remarkably insightful explanation considering the lack of knowledge in his time regarding the structure of matter. In Franklin's theory, a single electric fluid must permeate all matter. A positive charge is an excess of this fluid, while a negative charge is a deficit. With this theory, Franklin explained the observed conservation of electric charge—when a positive charge was created, an equal negative charge was also created. We now know that Franklin's proposed electric fluid consists of electrons. Franklin and other scientists of his day had no way of knowing that an excess of electrons results in a negative charge, while a deficit results in a positive charge. Which charge was positive and which negative was an arbitrary convention that was decided before electrons were identified as the primary carriers of electric current.

Because the force of gravity varies with the inverse square of distance, physicists suspected that the electric force should also obey an inverse square law. Encouraged to study electricity by Benjamin Franklin, British chemist Joseph Priestly (1733-1804) experimentally proved the inverse square law for the electric force around 1766. French physicist Charles Augustin de Coulomb (1736–1806) improved the accuracy of Priestly's measurements with his invention of the torsion balance in 1777. He confirmed the inverse square law and showed that the electric force between two charges is proportional to the product of the charges, another analogy with the Newton's formula for gravity (1.1). In modern notation, Coulomb's formula for the force between two charges in free space is

$$F = -\frac{q_1 q_2}{4\pi\varepsilon_0 r^2} \quad (1.2)$$

where q1 is one charge, q2 is the other, r is the distance between the charges, and e0 is a universal constant. The minus sign in (1.2) indicates that the force between two charges of the same sign is repulsive rather than attractive. He found that the attraction between two magnets obeys a similar law. Coulomb also demonstrated around 1780 that magnets and static electric charges do not interact.

Figure 1.2—Coulomb's Torsion Balance

We temporarily depart from the main course in our historical banquet to sample a side order of frogs' legs. In 1786, Luigi Galvani (1737–1798), professor of anatomy at the University of Bologna, was performing dissections of frogs when his assistant, Giovani Aldini, happened to touch the legs of a recently deceased frog with a metal scalpel near an electric generator. To his great surprise, Aldini saw the legs twitch. Galvani repeated the experiment and noted that the muscular contractions were most pronounced when the generator—similar to the one invented by von Guericke—was producing sparks. Galvani and Aldini launched a series of experiments and demonstrations subjecting the muscle tissues of various recently deceased animals to electric currents. Aldini's most macabre demonstration involved human tissue in the form of an executed murderer's corpse, whose legs and arms were made to move with electric currents after the convict had been dead for an hour. This incident helped inspire the famous novel *Frankenstein* by Mary Wollstonecraft Shelley (1797–1851).

Galvani's experiments, interesting as they were, ultimately did not have the lasting scientific significance they were assumed to have at the time. They did not lead to a basic understanding of the biological process of life, contrary to many expectations. Galvani missed an important clue about the nature of electricity when he completed an electric circuit to a frog's leg using two different metals and noted that the leg moved without an external generator. He incorrectly believed that the muscle tissue stored static electricity and the circuit allowed the electricity to be discharged. Galvani's lasting contribution to the science of electromagnetics was to excite the interest of contemporary Italian scientist Alessandro Volta, who found the correct explanation for the source of electric current in the absence of a mechanical generator.

Alessandro Volta (1745–1827)—another friend of Benjamin Franklin and a proponent of the single fluid theory of electricity—repeated Galvani's frogs' legs experiments, but with a difference: he used live frogs. He connected two different metals in a circuit to a live frog and noted muscle contractions. At first, like Galvani, he attributed the effect to a source of electric current in the frog's tissues. However, when he touched a circuit made of two metals to his tongue and noted an unpleasant taste, he guessed that the source of electricity was not in the tissue, but in the moist interface between two different metals. He experimented with many different arrangements of metals and interfaces for eight years, finally announcing his discovery to the public in 1800. His invention, now known as the *voltaic pile*, consisted of stack of *voltaic cells*. Each cell was a sandwich of silver and zinc separated by a piece of moist cardboard. Today's batteries are refinements of Volta's discovery. Volta also coined the phrases, *electromotive force* and *electric current*, still in common use.

Once Volta made his invention public, it was easily replicated in laboratories in Europe and America. For the first time, experimenters could generate reproducible levels of continuous electric current, making systematic laboratory experiments easier to perform and repeat. As bigger and more powerful piles were constructed, practical uses were found for electric current: extracting metals from ores, driving chemical reactions in liquid solutions, and sending messages over wires—*telegraphy*—to list a few examples. Telegraphy was the first economically significant technology that relied on the coupling of electricity and magnetism, which is the next topic in our history.

Figure 1.3—Voltaic Pile

Prior to 1820, many researchers in electricity and magnetism believed that the electric and magnetic forces were mathematically similar but fundamentally different, like gravity. Coulomb's experiments had detected no mutual forces between permanent magnets and static electric charges, lending strength to this belief. Hans Christian Oersted (1777–1851), influenced by the writings of German philosopher Immanuel Kant (1724–1804), believed that there must be a deeper unity between the two forces. There was also a tantalizing hint of the connection in the observation of sailors that a lightning strike on a ship's mast could sometimes reverse the polarity of its compass. In the winter of 1820, he prepared an experiment that he tried for the first time during a lecture to his unimpressed class at the University of Copenhagen. In this experiment, he detected a small deflection of a compass needle in the presence of an electric current. Because the effect was so small, he hesitated to draw immediate conclusions and waited three months before resuming the experiments. After further trials, he realized that a thicker wire would allow more current to flow and produce a larger effect. He concluded that the current must generate a magnetic force because only a magnetic force could deflect the compass needle. Oersted gave the name *electromagnetism* to the effect he discovered.

When Oersted's experiment was reported in July 1820, it was simple for others to duplicate his result. Why, then, had it taken twenty years from the invention of the voltaic pile to notice that electric currents generate magnetic forces? Part of the explanation is that the direction of the magnetic force surprised physicists. Prior to Oersted's experiment, those who believed that an electric current might create a magnetic force expected that such a force would cause a compass needle either to align with the direction of the current or to point toward the wire. Instead, the observed magnetic force was circulating around the wire. Once the circulating magnetic field was observed, efforts began almost immediately to understand the phenomenon and to put it to use.

Following Oersted's announcement, German scientist Johann Schweigger constructed a more sensitive detector consisting of multiple turns of wire wound around the case containing the compass needle. Schweigger called his device a *galvanic multiplier* and it was the first accurate means of measuring electric current. Schweigger's instrument was the basis of electro-mechanical voltage and current meters found today in such places as music amplifiers and automobile dashboards. A similar idea is also the basis of electric motors.

André-Marie Ampère (1775–1836) was in the audience on September 4, 1820, when François Arago (1786–1853) presented Oersted's results to the skeptical French Académie des Sciences. Had Coulomb not demonstrated that the electric and magnetic forces do not interact? The difference was that Oersted had detected a magnetic force due to electric current—charges in motion—rather than static electric charges. Ampère immediately began experiments in his own laboratory and within two weeks was able to confirm Oersted's observation and add some refinements of his own.

Ampère constructed his experiment so the effect of the earth's magnetic field would be canceled in the measurement region by counter-magnets. This enabled him to detect the effect with more reliability and sensitivity than Oersted's simpler arrangement. He also improved Schweigger's galvanic multiplier, giving it the present name, *galvanometer*. He then observed the fact that a compass needle points at right angles to the wire and changes direction (north-south) depending on whether it is placed above or below the wire. From these observations, he confirmed Oersted's assumption that the magnetic force circulates around the wire. He also deduced what we now call the *right-hand rule*—placing the thumb of the right hand along the wire, pointing in the direction of the current flow, the fingers of the right hand curl around the wire in the direction of the magnetic force. As we will see in a later chapter, we use right-handed coordinate systems when we solve electromagnetic problems so that the right-hand rule is enforced mathematically.

Using the same techniques that Coulomb used to measure electric forces, Ampère measured the magnetic forces between current-carrying wires. The simplest case involved a pair of parallel wires. He found that wires carrying currents in the same direction attract each other. Wires carrying currents in opposite directions repel each other. Using modern notation, under the assumptions that the wires are thin and long compared to the distance between them, the force on one wire due to the other is

$$F = +\frac{\mu_0 i_1 i_2}{2\pi r} L \qquad (1.3)$$

where i1 is the current in one wire, i2 is the current in the other, r is the distance between the wires, L is the length of the wires, and m0 is a universal constant. The explicit plus sign in (1.3) reminds us that the force between two currents of the same sign is attractive while the force between two currents of opposite sign is repulsive.

Ampère's next step was to reason that if straight currents generate circular magnetic fields, then circular currents should generate straight magnetic fields. He wound helical turns of wire around glass cylinders and noted that the magnetic forces generated were similar to those that would be created by cylindrical permanent magnets. He placed an iron rod inside the coil and it behaved like a permanent magnet when the current was turned on. Ampère made an intuitive leap to the hypothesis that magnetic fields in permanent magnets must be created by tiny circulating currents within the material. This hypothesis reduced magnetism to a property of electricity. The two forces were not just connected, they were actually two aspects of the same underlying phenomenon.

There was one big problem with Ampère's hypothesis: How could current circulate indefinitely, particularly in a material like magnetite that is not a bulk conductor of electric current? Ampère's colleague and friend, Augustin-Jean Fresnel (1788–1827), suggested that microscopic currents flow around molecules. If the currents could be coaxed into alignment, then the material would exhibit a net magnetic force. This theory explained Gilbert's findings that magnets could be created by hammer blows (coaxing the circulating currents into alignment) and demagnetized by heating (destroying the alignment). It had the added advantage of being difficult to disprove because the properties of molecules were unknown at the time.

We know today that magnetization of a material results principally from alignment of the spins (intrinsic magnetic moments) of electrons. Nevertheless, the hypothesis of microscopic circulating currents was sufficient to allow Ampère to calculate magnetic forces from permanent magnets. His results were in good agreement with Coulomb's measurements. At the macroscopic or bulk level, there is no difference between magnetic forces generated by spinning electrons and magnetic forces generated by hypothetical currents circulating around molecules. The incorrectness of Ampère and Fresnel's theory is evident only when we account for the details of how materials are magnetized.

Around the same time as his suggestion to Ampère, Fresnel made an important discovery about light that was used later to confirm that light is a type of electromagnetic wave. Fresnel was a proponent of the wave theory of light introduced by Dutch physicist Christiaan Huygens (1629–1695)—as opposed to the particle theory of Newton.

Scientists of Fresnel’s time called the medium in which light waves travel the *luminiferous ether*, although they were unsuccessful in determining the physical properties of luminiferous ether for reasons we will discuss later. Studying the interference patterns of polarized beams of light, Fresnel deduced that the oscillations of light waves must be *transverse* (at a right angle) to the direction of the wave. Ocean waves are an everyday example of transverse waves—water rises and falls vertically as the waves propagate horizontally. Prior to Fresnel’s finding, proponents of the wave theory of light believed that light waves, like sound waves, were longitudinal waves, meaning that the oscillations occur in the same direction the wave travels. We will see in a later chapter that Maxwell’s formulation of electromagnetic theory succinctly explains the transverse nature of light waves.

Volta’s invention of the battery had allowed repeatable experiments with electric currents from 1800 forward, but predictive mathematical models of electric circuits were unknown. Around 1827, Georg Ohm (1789–1854) developed a mathematical theory of circuits that is now recognized as fundamental to the design of electrical and electronic devices. Ohm taught physics and mathematics at the Realschule in Bamberg, Germany, and at the Jesuits’ College in Cologne. During his years at Jesuits’ College, 1817–1826, he experimented with electric currents flowing through wires of various diameters and lengths. His research budget was small and his experiments were crude by today’s standards. Nevertheless, he was able to deduce in 1827 the law that now bears his name:

$$I = \frac{V}{R} \qquad (1.4)$$

where I is the electric current flowing through a wire, V is the electromotive force (also called voltage or electric potential difference), and R is the resistance of the wire.

Stung by ridicule of his theories and of his ponderous style, he resigned from Jesuits’ College and spent the next six years as a private tutor. In 1827, he published a monograph, *The Galvanic Current Investigated Mathematically*, which established the principles for determining the distribution of current and voltage in a circuit. Initially, his work received little attention in the scientific community. Only later in his career did he receive recognition for his achievements. He was made director of the Polytechnic School of Nuremberg in 1833. In 1841, the Royal Society in London awarded him the Copley medal. He was appointed professor of physics at the University of Munich in 1849.

As Ampère advanced the theoretical understanding of electromagnetism and Ohm advanced the theoretical understanding of electric circuits, other investigators pursued research that led directly to practical applications. In 1824, William Sturgeon, an officer with the British Royal Artillery who experimented with electricity in his spare time, constructed the first U-shaped electromagnet by winding a few turns of bare copper wire around an iron horseshoe. In 1829, Joseph Henry, whose many accomplishments we will discuss next, improved the idea by using many turns of insulated copper wire around a thicker U-shaped iron core. Eventually, Henry built a magnet that could lift a ton of iron. Within a decade of Oersted’s discovery of electromagnetism, scientists had created magnetic forces several orders of magnitude (factors of ten) greater than the forces from magnetite and other naturally occurring sources. Just as importantly, these forces could be controlled by regulation of electric currents. For the first time, inventors could contemplate generating and controlling magnetic forces of sufficient strength to do useful work. The dream of Peter Peregrinus to use magnetism as a motive force was finally about to be realized.

As with Benjamin Franklin, no history of electromagnetics would be complete without mentioning the contributions of Joseph Henry (1797–1878). Henry began his investigations of electricity and magnetism in 1827 while he was an instructor at Albany Academy, a private school for boys in Albany, New York. In 1829, he combined the ideas of Schweigger and Sturgeon to create the powerful electromagnet we have already described. He then attempted to create an even more powerful electromagnet with more turns of copper wire. He discovered there was a limit to the force he could generate—after a certain number of turns, adding more turns actually decreased the magnetic force. The extra turns increased the resistance (or impedance) of the circuit, decreasing the current.

In order to determine the best way to get the most power from a given voltaic pile or battery, Henry built an electromagnet using several shorter windings of insulated copper wire rather than one long winding. He could then connect the windings in series, which amounted to the same thing as one long winding, or connect them in parallel. He tested both types of connections with two types of batteries. The first type, which he called a *quantity* battery, consisted of one cell with large plates capable of producing high currents at a low voltage. The second type, which he called an *intensity* battery, consisted of many smaller cells connected in series, capable of producing high voltages at low current levels. He found that connecting the coils in parallel generated the highest magnetic force with the high-current, low-voltage battery. He called this type of magnet a quantity magnet because it worked best with a quantity battery. Conversely, connecting the coils in series generated the highest magnetic force with an intensity battery. He called a series-connected magnet an intensity magnet. Henry's solution to the problem of making the most powerful magnet possible using a given battery became the basis of the modern practice in electrical engineering called *impedance matching*—adjusting the impedances of the load and source to match.

Figure 1.4—Henry's Multi-Winding Electromagnet

Henry's investigations included experiments to show how electromagnetism could be transmitted over long distances. In 1824, English scientist Peter Barlow had dampened enthusiasm for this idea by demonstrating that the electromagnetic effect diminished considerably when the source was connected to the electromagnet through 200 feet of wire. Barlow concluded that using electromagnetism to communicate over long distances was impractical. Henry confirmed Barlow's result with a single-cell quantity (high current, low voltage) battery. He then tried the same experiment with an intensity (low current, high voltage) battery. This time, the electromagnetic effect was not significantly diminished by transmission through a long wire. Henry published his findings in 1831. Years later, this publication became a critical piece of evidence in contentious patent litigation involving Samuel Morse, the putative inventor of the telegraph.

In November 1832, Henry moved from Albany to Princeton, New Jersey to become professor of natural philosophy at Princeton College (later Princeton University). He continued with his electromagnetic experiments at Princeton, developing an early version of the *relay* in 1835. This consisted of two parts. The first was an intensity battery and switch that was used to transmit a signal over a long distance and activate an intensity magnet, in this case from one building at Princeton to another. The second part was a quantity battery and quantity magnet that would be switched on or off by the action of the intensity magnet. With this arrangement, a small current over a long distance could control a large current acting locally at the far end. The principles of Morse's telegraph were adapted directly from Henry's relay.

Figure 1.4B—Henry's Electromechanical Relay

Henry also developed an early version of the electric motor. In his device, a rocker arm controlled two switches in separate circuits for two electromagnets. Each electromagnet pulled down one side of the arm when activated. When the right side of the arm was down, the switch controlling the right electromagnet turned off and the switch controlling the left electromagnet turned on. This caused the arm to rock back to the left side, where the process reversed, rocking the arm back and forth about 75 times a minute. Although the motor was impractical for doing useful work, it made an engaging lecture demonstration.

Figure 1.5—Henry’s Rocker Arm Motor

Henry shares with Michael Faraday (1791–1867) the discovery of the principle of magnetic induction. Magnetic induction is the generation of electric current by a changing magnetic field. Faraday receives most of the credit because he published first and developed a more complete conceptual framework for magnetic induction that set the stage for the achievements of Maxwell. Henry is credited by historians of science with the correct interpretation in 1832 of the phenomenon of self-inductance. Self-inductance is the generation of electromotive force in a circuit due to changes in the magnetic field created by electric current in the same circuit.

Until this point in our history, we have been careful to use the term *magnetic force* rather than *magnetic field*. Before Michael Faraday, scientists did not have a clear concept of the electric and magnetic fields as entities in their own right. Faraday shifted the focus of research in electromagnetic theory from the sources of electromagnetism to the electromagnetic fields that exist throughout space.

If any scientist in our history can be said to have humble beginnings, it was Faraday. He was born to a poor family in a lower-class section of London. He had little formal education. At the age of nineteen, he saw a voltaic pile in operation at a meeting of the City Philosophical Society and soon built one of his own. In 1812, he attended a series of lectures on chemistry given at the Royal Institution by Humphry Davy (1778–1829). He made notes of the lectures and sent them to Davy with a request to be his assistant. Fate intervened when Davy was injured in a chemical explosion and he invited Faraday to be his unpaid secretary during his recuperation. Fate intervened again a few months later in 1813 when a laboratory assistant at the Royal Institution was fired and Davy arranged for Faraday to take his place. Faraday conducted research and taught at the Royal Institution for the rest of his life.

From 1813 to 1820, Faraday assisted Davy in his chemical experiments and had little opportunity to pursue his own experiments in electricity. However, Davy introduced him to many of the leading scientists of the time and Faraday continued to follow the progress of research in electricity and magnetism. In 1821, Richard Phillips, an editor of the *Annals of Philosophy* and a friend of Faraday’s, suggested that Faraday write an article summarizing what was known to that point about electromagnetism. Thus began the research that would occupy Faraday for the remainder of his career.

Faraday’s lack of formal education was in one way an advantage when he began to unravel the mystery of Oersted’s discovery—he had no preconceptions of what the magnetic field generated by an electric current *should* do. If the magnetic lines of force traced circular paths around a wire carrying current, he accepted that fact and allowed it to lead him to the implications. Not one to be content analyzing the experimental results of others, Faraday went to his laboratory and verified Oersted’s hypothesis of circular magnetic lines of force.

In order to understand what Oersted meant about circular magnetic forces, Faraday’s mentor, Davy, performed a simple experiment in which he passed a vertical current-carrying wire through a horizontal piece of paper and sprinkled iron filings on the paper. The iron filings formed circular patterns around the wire. Faraday saw this experiment and began similar ones of his own, placing permanent magnets in different configurations beneath paper and observing the patterns of iron filings. Many scientists had performed experiments with iron filings near magnets, but Faraday saw what they had not. In Faraday’s mind, the lines of magnetic force took on a reality of their own, independent of the iron filings. The old concept of action at a distance was replaced by the notion of fields filling all space, acting locally on objects intercepting the fields.

Figure 1.6A—Faraday's Iron Filing Patterns

Faraday considered the implications of circular magnetic lines of force around current in a wire. If he had a *magnetic monopole*—a north pole without a corresponding south pole (or south without north)—the magnetic field would force it to travel around the wire indefinitely as long as electric current flowed in the wire. Faraday did not have a magnetic monopole because, as far as we know, they do not exist. However, he conceived a clever experiment in which he placed a bar magnet in a beaker of mercury so that its south pole was fastened to the bottom and allowed to pivot, while the north pole was free to rotate in the beaker. He introduced a current through a wire touching the top surface of the mercury. The current continued down through the mercury and out an electrical connection at the bottom of the beaker.

If Faraday had switched on the current at this step, the demonstration would have worked. However, Faraday had a second point to prove. He understood that if a wire exerts a force on a magnet, then the magnet should exert an opposite force on the wire. He prepared another beaker of mercury in which a magnet was fixed vertically at the center and the wire coming from the top was free to rotate around the magnet, maintaining contact with the mercury. On Christmas Day, 1821, he connected both parts of his experiment in series and delighted to find the magnet in the first beaker and the wire in the second beaker rotating about their pivots, as he expected. Through his insight, Faraday had demonstrated the first continuous motion created by electromagnetism.

Figure 1.6—Faraday’s Demonstration of Circular Forces

The relationship of electric currents and magnetic fields puzzled Faraday for a decade, from 1821 to 1831. If an electric current could produce a magnetic field, there must be a way for a magnetic field to produce an electric current. However, all attempts to do so had failed. In 1831, Faraday learned that Joseph Henry was experimenting with reversing the current to an energized electromagnet. That hint was enough to inspire Faraday to try a new idea. He procured a ring of iron and wound it with two coils of wire on opposite sides. The first coil was connected to a switch and battery. The second coil was simply short-circuited (connected to itself) with a compass needle placed near the connecting leg to detect current (if any). His reasoning was that when the current was switched on for the first coil, a magnetic field would be created in the iron, which in turn would cause continuous current to flow in the second coil.

When Faraday tried the experiment, it was at first another disappointment—continuous current did not flow in the second coil. As he monitored the compass to check for a deflection, he noticed something curious. For a short moment when the first coil was switched on, the compass needle moved. Then it went back to its original orientation.

When Faraday switched the first coil off, the needle briefly moved again, this time in the other direction.

Other scientists should have and probably did notice this effect before Faraday. Only Faraday and Henry were mentally prepared to draw the correct conclusion: *changing* magnetic fields generate electric currents, not *static* magnetic fields. Faraday pictured magnetic lines of force crossing the second coil, inducing electric current to flow. When the magnetic field was established and no longer changing, the lines stopped crossing the second coil and current stopped flowing. Switching off the magnetic field caused the same effect in reverse. He published his results in 1832.

The missing element in previous attempts to find the link between electricity and magnetism had been *time*. Scientists from Thales forward had assumed that electric and magnetic effects were instantaneous. It now appeared to Faraday that electromagnetic effects could not propagate instantaneously; there must be a finite speed at which electric and magnetic fields move into space. However, Faraday did not know what that speed might be.

The mathematical expression of Faraday's discovery is called *Faraday's Law*. It states that the voltage induced in a loop of wire enclosing a changing magnetic field is proportional to the rate of change of the enclosed field. Faraday's Law was one of the building blocks used by Maxwell to formulate a complete theory of electromagnetics, as we will discuss below.

In addition to his discovery of magnetic induction, Faraday was the first to classify three types of materials with respect to their interaction with magnetic fields. *Paramagnetic* materials are weakly attracted to regions with higher concentrations of magnetic lines of force. *Diamagnetic* materials are weakly repelled from regions with higher concentrations of magnetic lines of force. *Ferromagnetic* materials are strongly attracted to regions with higher concentrations of magnetic force, and retain magnetic fields of their own in the absence of external sources. We will discuss these properties in more detail later in this book.

In 1845, Faraday discovered an effect that now bears his name (not to be confused with the *Law* that also bears his name). A young physicist named William Thompson (1824–1907)—who would later become Lord Kelvin after being recognized by the British Crown for his many contributions to science—wrote a letter to Faraday claiming that he had developed a mathematical basis for the notion of fields and that he suspected electric or magnetic fields should influence light propagating through an appropriate medium. Faraday failed to find a relationship between electric fields and light, but he found that a certain orientation of magnetic field altered the polarization of light passing through lead glass. This phenomenon is now known as the *Faraday Effect*.

Although Faraday had shown in 1845 that light could be influenced by a magnetic field, exactly what light is and in what medium it propagates were still mysteries. The first step to answering these questions was to find the finite speed at which magnetic and electric fields propagate into space as Faraday’s intuition had guessed. In creating his mathematical theory of electricity and magnetism, James Clerk Maxwell (1831–1879) would not only show how electromagnetic fields propagate, he would also show that the experiments needed to determine the propagation speed had already been performed by Coulomb and Ampère before induction was discovered.

James Clerk Maxwell was the scientific counterpoint to Michael Faraday. Maxwell came from a relatively affluent family in Scotland and he graduated from Cambridge University. Where Faraday was experimental and intuitive, Maxwell was theoretical and analytical. In short, they were the perfect complementary pair to establish a solid foundation for our present understanding of electromagnetics.

Maxwell began his research into the nature of electromagnetism in 1856 while a professor at the University of Aberdeen. To begin, he studied the results of the scientists who had gone before. While all of the results were important, we will mention three as the most important to the ultimate mathematical formulation that he developed.

First, Maxwell would base his theory on Faraday's idea of electric and magnetic fields acting locally. The relationships between those fields would be governed by equations in differential calculus much as Newton had formulated his laws of motion.

Second, Maxwell needed to explain how current flowing in a straight line could generate a circular magnetic field. This is not as simple as it may seem. For example, it was relatively easy for physicists to imagine straight lines of electric force emanating (*diverging*) from an electric charge—that was similar to the way they imagined gravity worked. But how could a line of force be created if it did not begin at a source?

Third, William Thompson had already solved part of the puzzle. Thompson had shown that the electric field generated by a set of static electric charges could be found using the same mathematical equations that govern the flow of heat in solid objects. This meant that Maxwell's theory in the case of static charges should reduce to the same mathematical form as the equations of heat. Maxwell appreciated the value of this important scientific hint and wrote to Thompson that he would "borrow it for a season."

It took nearly two decades, from 1856 to 1873, for Maxwell to develop his theory to its final form, which is essentially the same mathematical model that we use today. His theory is mathematically elegant, even beautiful. Much of this book will be devoted to finding solutions to Maxwell's equations for various applications. Before rushing into the mathematics of the theory, however, we will carefully lay the groundwork piece by piece, so that our physical insights are not lost in a maze of abstractions.

The fundamental concepts in Maxwell's theory can be reduced to three principles stated without mathematical symbols: First, *fluxes* (flows) are associated with electric and magnetic fields. We will defer the discussion of exactly *what* is flowing to a later chapter. For now, it suffices to know that wherever there is an electric field, an electric flux (flow) is also present, generally (but not always) pointing in the same direction as the electric field. Wherever there is a magnetic field, a magnetic flux is also present, generally (but not always) pointing in the same direction as the magnetic field.

Second, electric flux *diverges* from electric charges, while magnetic flux is *free of divergence* in the absence of magnetic monopoles. This means that electric charges are the beginning and ending points for electric flux. Magnetic flux, on the other hand, does not have a beginning or end.

Third, changing electric flux and/or moving electric charges create a *rotation* or circulation in the magnetic field, while changing magnetic flux creates a rotation or circulation in the electric field.

Maxwell expressed the concepts in his theory as equations in differential calculus using mathematics developed by William Rowan Hamilton (1805–1865) and William Thompson for general field theory. In Maxwell's day, the mathematical ideas behind the theory were considered so advanced that only the foremost theoreticians could understand it. Faraday, for example, was unable to penetrate the mathematics of Maxwell's theory although he was responsible for the physical insights on which it was based. Today, we consider understanding Maxwell's equations to be a necessary part of the undergraduate education of all scientists and engineers. Don't worry—we have learned a great deal since the time of Maxwell and Faraday about how to understand Maxwell's concepts and explain them with clarity and precision.

Maxwell's 1873 publication of *A Treatise on Electricity and Magnetism* marked the culmination of classical physics. The theory was an immediate success at Cambridge University, where Maxwell had been appointed the first Professor of Experimental Physics at the Cavendish Laboratory in 1871. Other scientists, particularly the French, were more skeptical. With any physical theory, it is not enough simply to explain the known phenomena. The true test of a theory is its ability to predict phenomena not yet discovered. In Maxwell's case, the true test was the prediction of electromagnetic waves.

Once Maxwell's equations were derived, the obvious next step was to find solutions to them. One extremely interesting set of solutions involved coupled electric and magnetic fields propagating in space and time. These are *transverse* waves in the sense we have already described. An additional interesting feature of these waves is that the electric and magnetic fields are transverse (at right angles) to each other as well as both being transverse to the direction of propagation. Maxwell found that the speed of propagation for these waves in vacuum should be

$$s = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \qquad (1.5)$$

where s is the propagation speed; m0 and e0 are the same constants we saw in (1.2) and (1.3) relating static forces to electric charges and currents. This was quite a remarkable result, because it predicted the existence of waves that had not yet been detected (as far as anyone knew) and it also predicted their speed using constants that were calculated from experiments in which nothing moved except electric current. When Maxwell first calculated the speed of electromagnetic waves using (1.5), the result was close to the best value then available for the speed of light, although the values for the speed of light, m0, and e0 that he used were not accurate by today's standards. As the accuracy of experimental results improved, it became evident that the value of (1.5) was exactly the speed of light. Whether the waves were in fact light, or existed at all, remained to be proven.

The electromagnetic waves predicted by Maxwell's equations matched two facts known about light: its speed and its transverse nature. However, that alone was not conclusive evidence that electromagnetic waves actually exist in the form predicted by Maxwell. Visible light was known to have a wavelength about 0.00005 centimeters (violet and blue a little shorter, yellow and red a little longer). Maxwell's theory predicted that electromagnetic waves should exist with any wavelength, not just wavelengths near visible light. Furthermore, no one had generated and detected an electromagnetic wave using an electrical source. Skepticism about Maxwell’s theory persisted until these objections were answered.

Heinrich Hertz’ discovery of electromagnetic waves in 1886 was a combination of good luck and the mental preparation necessary to take advantage of it. Hertz (1857–1894) became professor of physics at the Technische Hochschule of Karlsruhe, a small technical college in Germany, in March 1885. Plagued by self-doubt and loneliness, he questioned his ability to make a significant contribution to physics. He happened to find an induction coil in a cabinet and decided to conduct experiments with it. With this coil, he inadvertently built the first radio wave transmitter.

Induction coils were common in electrical laboratories of his time. An induction coil is similar in concept to the iron ring with two windings used by Faraday. The first coil, or *primary*, is connected to a battery with a switch. The second coil, or *secondary*, has many more windings than the primary, and so the voltage generated in the secondary circuit is much higher than the voltage of the battery. To make the most efficient use of the magnetic lines of force, both coils are wound together around a cylindrical iron core.

Hertz connected the secondary winding of the coil to two small metal spheres separated by a small air gap. When the primary coil was turned on (or off), an impressive spark jumped across the air gap of the secondary. The same basic principle is used today with automobile spark plugs. Hertz added larger metal spheres to the circuit with long metal rods so that he could vary the electrical characteristics of the circuit. In modern electrical engineering terms, he added electrical capacitance to the circuit and thereby created a *resonant circuit*—a circuit in which the current oscillates at a specific frequency. The large spheres and the metal rods acted as an antenna.

Figure 1.7—Hertz’s Electromagnetic Wave Generator

Although Hertz had inadvertently created a radio transmitter, serendipity was still needed in order to *detect* the radio waves he had created. His wife of less than a year, Elisabeth, happened to be with him in the laboratory on November 1, 1886, when one of them (some historians suggest that Elisabeth might have been the first) noticed that in addition to the main spark, smaller "side sparks" were occurring in unexpected places. The mystery of the side sparks piqued Hertz' curiosity.

Hertz soon built a detector circuit consisting of a small coil of wire and a spark gap that could be adjusted accurately. With the detector, he was able to map the strength of the electromagnetic signal in his laboratory. As expected, the signal was strongest close to the source. That much could be predicted from simple magnetic induction without an electromagnetic wave. The surprising result was Hertz' discovery that the signal dropped to zero at a certain distance from his apparatus, and then increased again as he continued to move away from the apparatus. This is the essence of wave behavior, convincing proof that Hertz had created and detected the first electromagnetic wave generated by an electric circuit. Hertz would later show that the waves he created could be refracted (bent) by passage through materials, just as light could. There was no longer doubt that Maxwell's equations correctly modeled the behavior of electromagnetic fields.

There is some controversy among historians of science about whether Hertz was expecting to find electromagnetic waves when he began his experiments and was only confirming his expectations, or whether his results were a total surprise and due principally to good luck. There is also controversy about whether Maxwell expected to find waves as a consequence of his theory. It is clear that Maxwell did expect waves, because he wrote in 1862 that

> We can scarcely avoid the inference that *light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena.* [Maxwell's italics]

Hertz was well acquainted with Maxwell's theory and he would have known that light was strongly suspected to be an electromagnetic wave. The controversy arises because neither Maxwell, Hertz, nor anyone else knew prior to 1886 how to detect electromagnetic waves with wavelengths on the order of meters—far longer than the wavelengths characteristic of light—and how to generate them directly from an electromagnetic source. There were mathematical clues in Maxwell's equations, but there was no experimental proof and no clear idea how one would even begin such experiments. The best answer to the controversy is to invoke the old adage: Fortune favors the well-prepared.

Hertz' validation of Maxwell's theory is a logical place to end the Scientific Revolution. The revolution began with three crucial developments: the idea that observations and experiments should provide the basis for theory, improvements in measurement techniques to support quantitative science, and the advanced mathematics needed for quantitative theories. What we now call *classical physics* is the sum of all the experiments and theories of the Scientific Revolution. Classical physics encompasses theories of mechanics, gravity, heat, optics, chemistry, and electromagnetics. Inventors successfully used these theories to create many devices we take for granted today: the telegraph, the telephone, radio, electric generators, electric motors, the incandescent light bulb, the portable camera, and the internal combustion engine, to name just a few. As the nineteenth century ended, many physicists lamented that all of the fundamental properties of nature were discovered and, like Alexander the Great, they had no more worlds to conquer. Like Alexander, they were wrong.

The Modern Era

Physicists at the beginning of the twentieth century believed that electromagnetic theory was essentially complete with the work of Maxwell. However, fundamental questions remained to be answered regarding the interaction of electromagnetic fields with matter. In addition, extensions of the theory and understanding of its deeper implications played a critical role in the development of ideas in modern physics in the twentieth century.

Conventional wisdom at the end of the nineteenth century held that light waves must travel through some kind of physical medium, called the *luminiferous ether* or just *ether* for short. In a sense, measurements of m0 and e0 indirectly measure properties of the ether—assuming ether exists in the first place. Despite the conviction that it must exist, direct attempts before 1887 to discover its physical characteristics had failed. Ether was hypothesized to fill all space and to define a universal fixed frame of reference. According to the ether theory, it should be possible to detect the motion of the earth with respect to ether's universal frame of reference because light on earth would appear to move faster in one direction than another, just as an approaching train appears to be moving faster to an observer in a train moving toward it than it would to an observer on the ground.

In 1887, Albert Michelson (1852–1931) and colleague Edward Williams Morley set out to demonstrate the existence of ether by detecting the motion of the earth with respect to it. The chief difficulty they faced was their assumption that the velocity of the earth (and all other astronomical objects) with respect to the ether must be very small compared with the speed of light. This assumption prevented them from simply measuring the speed of light in different directions, because any differences due to the earth's motion would be lost in measurement errors. Instead, they used Michelson's recent invention, the *interferometer*, to detect small differences in the speed of light in directions at right angles to each other. When they tried the experiment, they detected no differences despite the sensitivity of the interferometer.

Figure 1.8—Michelson's Interferometer

The Michelson-Morley experiment is the most famous null result in physics. It raised fundamental questions, not only about light and electromagnetics, but about Newton's laws of motion. In order to explain the null result of the Michelson-Morley experiment, Hendrik Antoon Lorentz (1853–1928) and George Francis FitzGerald hypothesized that material objects contract in the direction of motion with respect to the ether by a factor that is exactly the amount needed to explain the equal transit times of light waves in all directions. This effect is known as the *Lorentz-FitzGerald contraction*. Lorentz and FitzGerald came close to the correct explanation, but it was Albert Einstein who finally resolved the apparent paradox and revolutionized physics in the process.

According to his teachers, Albert Einstein (1879–1955) was not destined for success. Although possessed of a brilliant mind—he taught himself Euclidean geometry at the age of twelve—he hated the regimentation of formal education. He cut most of his classes, preferring to learn the material on his own by reading books and studying his classmates' notes. His professors at the Swiss Federal Polytechnic in Zürich would not recommend him for further academic training when he passed his undergraduate examinations in 1900. He worked for two years as a tutor and substitute teacher. In 1902, he was hired as a patent examiner in the Swiss patent office in Bern. It appeared that his teachers had been correct and Einstein was destined for an obscure life.

Meanwhile, Einstein's brilliant mind had not been idle. He used the free time his jobs allowed him to follow developments in physics and complete his doctoral dissertation at the University of Zürich. In 1905, he received his doctorate and published three papers (now "classics" of physics) that established his reputation. The first paper described a mathematical model of random motions of particles in a fluid, and is not relevant to our book.

Einstein's second paper concerned the puzzle of the *photoelectric effect*. Physicists had established by 1905 that electrons are responsible for electric current. In further confirmation of the electromagnetic nature of light, physicists had also demonstrated that light waves incident on a metal surface could transfer electromagnetic energy to the electrons, ejecting electrons from the surface and causing electric current to flow. The ejected electrons were called *photoelectrons*. The puzzle was that the photoelectric effect was observed only for incident light with a frequency above a critical value, called the *cut-off frequency*. If the incident light had a frequency below the cut-off frequency, no electric current flowed no matter how intense the light.

The sharp cut-off observed in the photoelectric effect puzzled classical physicists because they assumed that all physical processes should be continuous. Light with lower frequencies (longer wavelengths) might be less efficient at exciting the electrons than light with higher frequencies (shorter wavelengths), but zero electric current for lower frequencies was impossible to explain with classical physics. Classical physics also could not explain the experimental observation that the energy of the photoelectrons depended on the frequency of the light but not on the intensity of the light.

Einstein's explanation was both simple and extraordinary. He suggested that electromagnetic waves consist of tiny, indivisible bundles of energy. About 20 years later, when Einstein's explanation had been well established by experiments, the indivisible bundles of electromagnetic energy were called *photons*. The energy of a photon is related to its wavelength and frequency by the formula

$$\mathscr{E} = \mathrm{h}\,\mathrm{f} = \frac{\mathrm{h}\,\mathrm{c}}{\lambda} \qquad (1.6)$$

where E is the energy, c is the speed of light, l is the wavelength, f is the frequency, and h is a universal constant known as Planck’s constant.

The electrons in the metal are bound to the lattice of metal atoms by the positive charges of atomic nuclei. In order to leave the metal surface and conduct electricity, an electron must receive enough energy from the impact of a photon to overcome its binding energy, in the same sense that a rocket leaving the surface of a planet must have enough kinetic energy to escape the gravitational binding energy of the planet. According to Einstein’s explanation of the photoelectric effect, if a photon does not have enough energy—if its frequency is too low—the electron it impacts cannot escape from the metal surface. Two or more photons of lower energy cannot combine to free one electron because the probability of multiple photons hitting the same electron is zero for all practical purposes.

Although Einstein’s theory neatly (and correctly) explained the photoelectric effect, it was controversial. Fresnel and other researchers in the field of optics had proven conclusively that light is a wave. Einstein was going back to Newton’s concept of light as a stream of particles. How could light be both? Nothing in Maxwell’s equations allowed for the existence of indivisible bundles—*quanta*—of electromagnetic energy. Almost half a century would pass before the quantum theory of light was reconciled with Maxwell’s equations.

While Einstein’s explanation of the photoelectric effect was merely controversial, his third 1905 paper, "On the Electrodynamics of Moving Bodies," was revolutionary. It undermined a foundation of classical physics—the assumption that time is a universal reference ticking along at the same rate everywhere. This hypothesis seemed so self-evident before 1905 that it hardly needed stating. It stood as an implicit assumption behind every time-dependent equation of classical physics, including Newton’s laws of motion and Maxwell’s equations. In his paper, Einstein claimed that this assumption must be wrong.

Einstein’s paper contains two postulates. First, he postulated that the laws of physics are the same in any inertial frame of reference. We can think of an inertial frame of reference as an imaginary laboratory that is not accelerating. Our imaginary laboratory contains all the equipment we need to measure any event’s instant in time and position in three-dimensional space. Einstein’s first assumption states that there is no physical measurement we could perform that would establish a preference for one inertial frame of reference over another. Any two inertial frames of reference may be in motion with respect to each other, but there is no universal standard frame of reference that can be used to establish an absolute velocity.

Einstein’s second postulate is that the speed of light (or any electromagnetic wave) in vacuum is the same for any observer in any inertial frame of reference, regardless of how the light was created. This assumption was confirmed by the Michelson-Morley experiment. Nevertheless, it goes against our intuitive understanding of how velocities should add. For example, consider someone firing a rifle from a moving train at a target on the ground. Neglecting air resistance, we expect that if he fires forward, the speed of the bullet as it hits the target should be the sum of the muzzle speed of the bullet plus the speed of the train. If he fires backward, the speed of the bullet as it hits the target should be the muzzle speed minus the speed of the train. Einstein’s second postulate claims that if the shooter were firing a laser beam instead of a rifle, the speed of the laser beam as it hit the target would be exactly the speed of light no matter which direction he fired or how fast the train was moving, even if the train was moving at close to the speed of light itself.

Figure 1.9—Firing a Rifle from a Moving Train

Einstein applied these postulates to the case of a charged body moving in an electromagnetic field. In 1905, physicists knew that a charged body moving in a magnetic field experiences a force at right angles to both its velocity and the magnetic field. We call this force the *Lorentz force*. It is responsible for distortions of a TV or computer CRT display in the presence of a magnetic field, for example, as the electrons traveling in the vacuum tube are curved away from their original trajectories. For simplicity, we will assume in our imaginary experiment—what Einstein called a *gedanken* or *thought experiment*—that we have an electron moving in a magnetic field in our laboratory but there is no electric field. Now consider the situation from the frame of reference of the moving electron, that is, the frame of reference in which the electron has no velocity. According to Einstein's first postulate, physics should be the same in the electron's frame. If the electron experiences a force in one inertial frame of reference, it should experience a force in all frames. The only two possible electromagnetic forces on the electron are the electric force from an electric field and the Lorentz force from a magnetic field. Since by definition the electron is at rest in its own frame of reference, it cannot experience a Lorentz force in its own frame. Therefore, the force it experiences can only be due to an electric field as measured in the electron's frame of reference, although there is no electric field in our laboratory frame of reference. From this argument, it is clear that the components of the electric and magnetic fields can change depending on which frame of reference we use to measure them.

Einstein set out to derive a consistent set of equations governing the transformation of electric and magnetic fields from one inertial frame to another. By invoking the second postulate, the invariance of the speed of light, Einstein showed that only one transformation was possible. This transformation involved not only the electric and magnetic fields, but the space and time coordinates as well. The transformation of space and time coordinates—*space-time*—between moving inertial frames is the essence of Einstein's *special theory of relativity*. It was not surprising that the transformation involved space coordinates—the electron's space coordinates in its own frame of reference never change, while its space coordinates in our imaginary laboratory frame of reference constantly change. The surprising result was that *time* coordinates transformed as well. According to Einstein, a clock in a moving reference frame should seem to tick more slowly when measured in our imaginary laboratory. This effect, related to Lorentz-FitzGerald contraction, is called *time dilation*. The greater the speed of the moving reference frame, the slower a moving clock ticks. A clock moving at the speed of light (if it could) would not tick at all as measured in our laboratory.

To make matters more confusing for classical physicists, the time dilation effect works just as well the other way around—from the electron's frame of reference, clocks in our imaginary laboratory seem to tick more slowly. This is the essence of the famous *twin paradox* of special relativity. Imagine a set of twins, twin A who stays on earth and twin B who leaves on a fast rocket. According to twin A, twin B ages more slowly. According to twin B, from whose frame of reference twin A and the earth are traveling in the opposite direction, twin A ages more slowly. When B returns, is A or B younger? The solution to this paradox is found in Einstein's *general theory of relativity*.

Einstein published his general theory of relativity in 1916. In order to account for accelerating frames of reference, he added a third postulate: the effect of acceleration (change of velocity) on a frame of reference is locally indistinguishable from the force of gravity. The usual gedanken experiment physicists invoke to explain this postulate is to compare two imaginary closed elevator cars, one on the surface of the earth and one in space. If the car on earth is sitting still on the surface, an observer inside will notice a force pulling everything down toward the floor. The force on an object equals the mass of the object times 9.8 meters per second per second. The observer calls the force gravity. Now imagine the second elevator car in space, far from the gravitational force of any astronomical object, accelerating in the local "up" direction at 9.8 meters per second per second. Einstein's third postulate holds that there is no experiment an observer inside the elevator car can perform to distinguish the first case from the second.

Figure 1.10—Imaginary Elevator Cars

Using the three postulates of relativity, Einstein derived equations that describe the curvature of space-time in the presence of mass and/or energy, called *mass-energy* for short. The greater the concentration of mass-energy, the greater the curvature of space-time. The local curvature of space-time is responsible for the gravitational force. Einstein replaced the concept of gravitational force at a distance with the local curvature of space-time, in much the same spirit that Faraday and Maxwell replaced the concept of electric and magnetic forces at a distance with electric and magnetic fields acting locally.

Like Maxwell's electromagnetic equations, Einstein's equations contained solutions that could be tested experimentally. One difference between Newton's theory and Einstein's was the effect of gravity on photons, particles of light that have energy but no mass. According to Newton's law of gravitational force (1.1), photons should not be affected by gravity at all since a photon's mass is exactly zero. Einstein's theory predicted that the local curvature of space-time should curve the path of a photon. The first evidence of the correctness of Einstein's theory of general relativity was obtained in 1919, when observations during an eclipse of the sun confirmed that starlight bends when it passes near a massive object (the sun). Einstein's equations also explained small variations in the orbits of planets that depart from Kepler's and Newton's laws of motion.

All experiments and observations since 1919 designed to test Einstein's general theory of relativity have supported it, or at least failed to disprove it. One of the stranger predictions of Einstein's equations was the possible existence of objects so dense that the local curvature of space-time would prevent light (or anything else) from escaping—*black holes*. By 1973, astronomical x-ray sources were observed that matched predictions of the radiation that would be emitted by matter falling into a black hole. In 1994, the Hubble Space Telescope provided evidence that a huge black hole exists at the center of galaxy M87. Our own galaxy may have a black hole at its center as well.

A critical prediction of Einstein's equations not yet proven or disproven is the existence of gravitational waves. Gravitational waves in theory consist of fluctuations in the curvature of space-time that travel at the speed of light. They are created by accelerating masses, just as electromagnetic waves are created by accelerating electric charges. One difficulty in designing gravitational wave detectors is that, in the absence of detected waves, physicists can only estimate the required sensitivity based on hypothetical catastrophic gravitational events such as the collision of two black holes. Only an extremely massive object undergoing extreme acceleration generates sufficient gravitational wave energy to be detected over astronomical distances. Even assuming events of sufficient violence, detecting a gravitational wave on earth requires a complex cryogenic system and support structure so that thermal and seismic noise sources are minimized. Instruments designed to detect gravitational waves have so far not been sensitive enough to detect any that might exist near the earth.

As this book was being written, the most sensitive detector yet, the Laser Interferometer Gravitational-Wave Observatory — LIGO — was under construction for a proposed commissioning in 2001. It consists of two virtually identical facilities, one in Hanford, Washington, and the other in Livingston, Louisiana. By comparing the signals from both locations, extraneous noise signals are more easily identified. Also, again assuming gravitational waves are discovered, differences in arrival times at the two sites will help locate the gravitational wave source, which would ideally correlate with an object that also emits electromagnetic waves. The existence or non-existence of gravitational waves will likely be resolved by LIGO in the next few years.

The confirmation of Einstein’s general theory of relativity in 1919 established his fame as the greatest physicist of his time. In his later career, he pursued an unsuccessful search for a theory unifying gravity with the other fundamental forces. The search for such a theory continues today.

Once Einstein’s general theory of relativity was complete, the solution to the twin paradox was straightforward. The situations of twins A and B are not completely symmetric. Twin B’s rocket had to accelerate away from earth, accelerate to turn around to return to earth, and finally accelerate to land on earth again. By Einstein’s third postulate, twin A was accelerating the whole time, but in a more subtle way, by experiencing the force of gravity on earth. If both twins use Einstein’s general theory of relativity to calculate their ages at the end of B’s journey, they will agree that B is slightly younger than A because B’s path through space-time experienced more time dilation than A’s. An equivalent experiment involving two identical atomic clocks, one remaining in place while the other was flown around the world, confirmed the predictions of Einstein's theory.

After Einstein’s special and general theories were confirmed, the hypothesis of luminiferous ether was quickly dropped. The medium in which electromagnetic waves propagate is free space (vacuum)—in other words, nothing at all! It is easier to imagine particles traversing free space than waves. This brings us back to Einstein’s explanation of the photoelectric effect. How can the concept of particles of light—photons—be reconciled with Maxwell’s equations that describe light as a continuous wave of electric and magnetic fields?

To understand how photons can be particles as well as waves, we will first consider the discovery of electrons and how they were shown to be waves as well as particles. Scientists from Du Fay and Franklin’s time forward had postulated that electric currents must be carried by some type (or types) of electric fluid, but the nature of the fluid was elusive.

English physicist William Crookes (1832–1919) made possible the study of electric currents in the absence of a conducting wire with his invention of the *Crookes’ tube* or *cathode ray tube* around 1879. In its simplest form, a Crookes’ tube consists of a partially evacuated glass tube with a metal electrode at either end. One electrode is negative (the *cathode*) and the other is positive (the *anode*). If the potential difference (voltage) between the anode and cathode is sufficient, not only will current flow in the tube, but certain regions of the residual gas in the tube will glow. This effect is the basis of today’s fluorescent lights.

Figure 1.11—Crookes' Tube

Crookes and other scientists soon learned that the current could be increased dramatically by heating the cathode, but not the anode. In 1897, English physicist Joseph John Thomson (1856–1940) postulated that some type of negatively charged electric fluid, *cathode rays*, must emanate from the cathode. He identified cathode rays as *electron beams*—streams of the negatively charged components of atoms—in 1899 and estimated from considerations of heat transport in cathode ray tubes that electrons must be about 1,000 times lighter than hydrogen ions (protons). By measuring the deflection of an electron beam in a magnetic field, he was able to calculate the ratio of charge to mass for electrons accurately, but not the mass and charge separately. According to Newton's and Maxwell's laws, if an electron had twice the charge it would be subjected to twice the force from a magnetic field, but if it had twice the mass its deflection in the magnetic field would be the same.

American physicist Robert Andrews Millikan (1868–1953) realized that if either the electron's mass or charge could be measured accurately, the other quantity could be calculated using the electron's known charge-to-mass ratio. In 1909, he conducted his famous oil-drop experiment at the University of Chicago. By observing the acceleration of oil droplets in a known electric field between two charged plates, he found that the charges on the droplets occurred in small multiples of -1.60×10^{-19} coulomb, the charge of a single electron. This experiment confirmed that the carriers of electric charge are quantized—electric charge cannot be subdivided into arbitrarily small amounts. In another experiment in 1916, Millikan confirmed Einstein's equation for the photoelectric effect (1.6) and measured an accurate value for Planck's constant, h, which plays a critical role in setting the limits of classical physics.

The experimental work of Millikan and others in the early part of the twentieth century led to the inescapable conclusion that the behavior of matter on a microscopic scale—atom by atom, electron by electron—is governed by the laws of *quantum mechanics*, not classical physics. While the subject of quantum mechanics is beyond the scope of our book, the two fundamental principles of quantum mechanics bear on our discussion of electromagnetics and its application to matter on an atomic scale. These two principles are *uncertainty* and *complementarity*.

German physicist Werner Heisenberg (1901–1976) developed the uncertainty principle in 1927. It states that certain pairs of variables describing the state of a physical system cannot simultaneously be measured with arbitrary accuracy. The product of the uncertainties in these pairs of variables must equal or exceed Planck's constant divided by 2p. For example, the uncertainty in the location of an electron, which we will call Dx, and the uncertainty in its momentum, which we will call Dp, obey the inequality

$$\Delta x \, \Delta p \geq \frac{h}{2\pi} \qquad (1.7)$$

Danish physicist Niels Henrik David Bohr (1885–1962) extended the idea of uncertainty by introducing the principle of complementarity in 1928. His principle states that for any *dynamical system*—a physical object or process that evolves with time—a complete description of its state requires at least one complementary pair of variables that obey the uncertainty principle. Because the value of Planck's constant is very small (about 6.626×10^{-34}joule-sec), uncertainty normally plays an insignificant role in the dynamics of large systems. However, the principles of uncertainty and complementarity prevent classical physics from working at an atomic level, because the models of classical physics assume complete and exact knowledge of all variables of a dynamical system.

How are we to understand the uncertainty principle in relation to events in the world accessible to our senses? According to American physicist Richard Feynman, whose role in reconciling electromagnetics with quantum mechanics we will discuss shortly, we should simply accept that this is how physics works at the atomic level and move on. However, a more intuitive insight is possible based on a theory developed by French physicist Louis Victor, Prince de Broglie (1892–1987) in 1924. De Broglie's theory was one of the *eureka* moments in physics. While still a graduate student, he proposed a novel idea: any particle of matter can be considered a wave as well. The wavelength of a particle is given by

$$\lambda = \frac{h}{p} \qquad (1.8)$$

where p is the momentum of the particle. De Broglie's idea was confirmed in 1927 by British physicists George Paget Thomson (son of J.J. Thomson) and Clinton Davisson observing electron diffraction in crystals.

If particles are also waves, the uncertainty principle follows immediately. We call the mathematical function that describes a particle's wave nature its *wave function*, usually assigned the Greek letter Y in equations. Like electromagnetic waves, a wave function exists in vacuum. It represents how likely we are to find the particle at a given location and time. Consider an electron moving in the +x direction with a momentum that is known exactly. In that case, the wave function for the electron is

$$\Psi = \cos\left\{\frac{2\pi(x - v\,t)}{\lambda}\right\} \quad (1.9)$$

where v is the electron’s velocity in the +x direction. (For simplicity, we show only the real part of the wave function. Wave functions are complex functions having real and imaginary parts.) This function has no definite beginning or ending in space. In other words, the electron’s uncertainty in momentum is zero, but its uncertainty in location is infinite.

The waves in our everyday experience, like ocean waves, are *localized*—confined to a limited region of space. To localize a wave mathematically, it is necessary to superimpose many different wavelengths in one wave function. More precisely, a continuous spectrum of different wavelengths must be superimposed. Because many different wavelengths are superimposed, the wave function cannot be assigned a single value of wavelength (or momentum) but only an average value and a distribution around that value. The width of the distribution of wavelengths correlates to the uncertainty in momentum. As the wave function becomes narrower in the x direction, decreasing the uncertainty of its location, a wider spectrum of superimposed wavelengths is required, thus increasing the uncertainty of its momentum. At the extreme, a wave function that localizes the wave exactly requires superposition of every possible value of wavelength (or momentum) from zero to infinity—the uncertainty in the electron’s location is then zero, but the uncertainty in its momentum is infinite.

Figure 1.12—Localized Wave Function and Wavelength Spectrum

Recognition of the dual particle-wave nature of matter resolved the seeming paradox in Einstein’s explanation of the photoelectric effect. If an electron or an atom can be a wave, then it is not too difficult to imagine that an electromagnetic wave can also be a particle. Only when combining many atoms or many photons does the distinction between a particle and a wave emerge.

By the late 1920’s, it was evident to physicists that the new theories of relativity and quantum mechanics had passed all experimental tests. However, the basic formula of quantum mechanics, Schrödinger’s equation, was based on mathematical ideas from classical mechanics and did not take relativistic effects like time dilation into account. In 1928, P.A.M. Dirac (1902–1984) developed a new formulation of Schrödinger’s equation that incorporated the theory of relativity.

Dirac applied his equation first to the electron. His formula predicted the existence of negative energy states of the electron as well as positive energy states. To explain the fact that electrons do not decay into negative energy states, Dirac proposed that (nearly) all negative energy states of electrons are already filled. By the rules of quantum mechanics, two electrons cannot occupy the same quantum state. Therefore, positive energy electrons cannot fall into negative energy states. The negative energy electrons are normally not observable because they uniformly fill all space.

According to Dirac’s theory, if a negative energy electron happened to receive enough energy to kick it into a positive energy state, then what would be observed would the creation of an electron plus a hole in the sea of negative energy electrons, where before there was nothing. The hole would behave as if it were a particle of the opposite (positive) charge and the same mass as an electron. The electron and hole could also recombine, releasing energy.

In 1928, nothing resembling Dirac’s prediction had been observed in the laboratory. Thinking that something must be missing from his theory, Dirac attempted without success to assign the role of the positively charged holes to protons. Protons were at that time the only observed particles with positive charge. However, the mass of the proton was too large. Also, protons and electrons do not recombine into energy, but form stable hydrogen atoms.

Dirac should have had more faith in his equation (as he later admitted) because in 1933, Carl D. Anderson detected particles in cosmic rays that had the same mass as electrons but a positive charge, called anti-electrons or *positrons*. Furthermore, physicists also observed positrons and electrons mutually annihilating to release energy in the form of photons. Eventually, physicists discovered that all particles of matter have their own anti-particles, as Dirac’s equation predicts. If sufficient energy is present, or the time span involved is so short that the uncertainty principle allows sufficient energy to be "borrowed" from the vacuum, any particle and its anti-particle can be created at any point in space. This principle plays an essential role in the theories of modern physics.

Dirac added the principles of relativity to quantum mechanics, but the final reconciliation of electromagnetics with quantum mechanics remained incomplete until the late 1940’s. Enter Richard Feynman (1918–1988), the most charismatic and iconoclastic physicist of the twentieth century. Feynman was one of the youngest physicists to work on the *Manhattan Project*, the building of the first atomic bomb. He quickly gained a reputation for his sense of humor and disdain of authority as well as his scientific insights. Toward the end of his career, Feynman served on the committee investigating the space shuttle *Challenger* disaster. He came to the attention of the mass news media when, during a committee meeting, he cut short an extended explanation of the low-temperature behavior of *Challenger’s* failed rubber seal by dunking a sample into his glass of ice water and demonstrating its low-temperature behavior himself. Feynman comes into our history because of his development, along with colleague Murray Gell-Mann (1929–), of the theory of *quantum electrodynamics*, or QED for short.

In order to account for electromagnetic interactions at the quantum level, Feynman and Gell-Mann abandoned the field model of electromagnetics and switched to the particle or photon model. The major difficulty with using the photon model is that all computations involve the interactions of quantum wave functions governed by the uncertainty principle. Rather than the simple deterministic results of classical physics, calculations in quantum mechanics result only in probabilities.

QED is well beyond the scope of this book, but the basic principles behind QED explain the apparent paradox between the existence of photons and Maxwell's equations. To get an idea of how QED works, consider the interaction of two electrons traveling in free space. In classical physics, we would assign each electron an initial location and momentum (or velocity) in three-dimensional space at a starting point in time. Using Newton's laws of motion—modified by special relativity—and the formula for the electrostatic force (1.2), we would derive a coupled set of differential equations relating the position, momentum, and acceleration of both electrons. To make the calculation simpler, we would select the center of momentum of the electron pair as our frame of reference, since this frame of reference would be inertial and the combined momentum of both electrons would remain zero at all times. We might solve this initial value problem by closed-form techniques or numerical methods.

To complicate matters, there are two additional terms to consider: the magnetic dipole moments of the electrons and radiation. As we previously mentioned, all electrons have an intrinsic magnetic dipole moment, as if each electron were a tiny bar magnet. The magnetic moment is aligned with the electron's spin. While the mutual force generated between the electrons by their respective magnetic moments is small compared to the electrostatic force of repulsion, we must include the magnetic force if we want our calculation to be accurate. Therefore, to the electron's position and momentum we must add the orientation of its magnetic (or spin) moment to describe its state completely in the terms of classical physics.

The second additional term is radiation. When any charged particle accelerates, it radiates electromagnetic energy. This radiation exerts a force on each electron and converts a portion of their kinetic energy into electromagnetic wave energy. With the magnetic force and radiation terms included, we can complete our classical physics calculation and predict exactly the location, momentum, and magnetic moments of both electrons at any future time.

Unfortunately, our classical physics calculation is incorrect. It makes the following assumptions that violate the principles of quantum mechanics:

- The position and momentum of both electrons are specified exactly, violating the uncertainty principle.
- The spin directions of both electrons vary continuously, violating a principle of quantum mechanical spin.
- Electrons radiate electromagnetic energy in continuous increments instead of individual photons.

In order to bring the calculation into accord with the principles of quantum mechanics, QED drops the ideas of action at a distance and continuous electromagnetic fields, substituting the exchange of particles—photons—to transfer energy and momentum from one electron to the other and from one electron to the radiated photon(s). Unlike classical physics, the result of a calculation in QED is not the exact trajectories of the particles, but instead a set of probabilities that the particles will end in each possible quantum state after the interaction. To test the calculation, we would devise an experiment with two colliding beams of electrons. We would place electron detectors in many directions (or move one detector to take readings in many spots) and count the total electrons hitting each one in a given length of time. We might also install photon detectors to check the radiated electromagnetic energy. Once we had enough data to be statistically significant, we would compare our measurements with the calculated predictions of the numbers and energy distributions of electrons and photons in the detectors. If we performed the calculations and the experiment correctly, the results should agree within statistical error. However, it would be impossible for us to know for any particular pair of colliding electrons which detectors they would hit.

Calculation of the probability function is a complicated procedure that requires integration of wave functions, interaction functions, and other functions known as propagators. Feynman developed a pictorial method known as the *Feynman diagram* to assist in setting up the calculations of QED. A simple first-order diagram is show below in Figure 1.13. The direction of time is up. The electrons (straight lines) interact by exchanging one photon (wiggly line), which is absorbed at the end of the process.

Figure 1.13—First-Order Feynman Diagram

Higher-order interactions involving the exchange of more photons and the release of photons are also possible. As the interactions become more complex, their probabilities decrease. The calculations are cut off at a level of complexity corresponding to the required accuracy.

The calculations of QED are very different from the techniques used for calculations in classical physics. It is certainly not obvious where Maxwell's equations fit into QED. However, Feynman and Gell-Mann were able to show that QED reduces to Maxwell's equations in the case of large bodies interacting over long distances compared to the scale of an atom. Maxwell's equations describe the average behavior of large numbers of photons in the same way that the equations of thermodynamics describe the average behavior of large numbers of molecules. For most practical purposes in electromagnetics, that is the appropriate level of detail.

Conclusion

As we begin the twenty-first century, some physicists imagine that all of the fundamental principles of physics have been discovered and nothing remains but filling in the details. Does this sound familiar? Based on history, we can expect that there are more surprises in store. Perhaps LIGO will detect gravitational waves of unexpected intensity from an unexpected source, or cast doubt on Einstein’s theory of general relativity by failure to detect gravitational waves at all. Perhaps a new unified field theory will lead to deeper insights and predict astonishing new phenomena. In any case, physics has clearly advanced to the stage that new discoveries will likely require teams of experts and the investment of millions or billions of dollars. The days are long gone when fundamental research could be conducted on a tabletop.

In our history of electromagnetics, we have mentioned only some of the advances in pure and applied mathematics that have supported the development of electromagnetic theory. As we progress through the remainder of the book, we will include a brief historical note each time that we introduce a new mathematical technique. In the past, an intimate familiarity with the techniques of applied mathematics was essential to performing calculations in electromagnetics. While understanding the techniques for finding closed-form solutions to special cases remains important, calculations in electromagnetics now rely on numerical methods executed on high-speed digital computers.

We are only beginning to realize what we can accomplish by solving the existing equations of physics on high-speed digital computers. Microprocessors capable of performing billions of operations per second will be commonplace within the first decades of the twenty-first century. Coupled with large memories and fast network connections, the computers of the twenty-first century will allow physicists to solve the equations of electromagnetics, quantum mechanics, and general relativity for large and complex systems in three and four dimensions. For example, astrophysicists will be able to calculate with precision the formation of planets, the evolution of stars, and the dynamics of galaxies.

Closer to the subject of this book, engineers designing new products such as integrated circuits and personal communication devices will perform complete three-dimensional, time-dependent analyses coupling electromagnetic equations with mechanical, thermal, and solid-state equations. Further, these analyses will be embedded in automatic optimization algorithms that will tune the designs to achieve the performance and cost goals set by the designer. Broad knowledge of physics, mathematics, numerical methods, and the various engineering disciplines will be more important for most practicing engineers than specialization in one area of engineering.

Last update: July 04, 2000