Ministry of Education and Science of the Russian Federation
Federal State Budgetary Institution of Higher Education
UFA STATE AVIATION TECHNICAL UNIVERSITY

IGOR V. ALEXANDROV, DENIS Y. VASIL’EV

FUNDAMENTALS OF PHYSICS TEXTBOOK FOR LABORATORIES

Уфа 2021

Ministry of Education and Science of the Russian Federation
Federal State Budgetary Institution of Higher Education
UFA STATE AVIATION TECHNICAL UNIVERSITY

IGOR V. ALEXANDROV, DENIS Y. VASIL'EV

FUNDAMENTALS OF PHYSICS TEXTBOOK FOR LABORATORIES

Approved by Editorial and Publishing Council of USATU
for the bachelors students in Applied Aviation

Учебное электронное издание сетевого доступа

ISBN 978-5-4221-1450-4

Ufa 2021

Reviewers:
PhD, Professor, Gennady P. Mikhailov;
PhD, Associate Professor, Oleg K. Babkov

Alexandrov Igor V., Vasil'ev Denis Y.
Fundamentals of Physics Textbook for Laboratories : [Электронныи ресурс] / Уфимск. гос. авиац. техн. ун-т. – Уфа : УГАТУ, 2021. – URL:
https://www.ugatu.su/media/uploads/MainSite/Ob%20universitete/Izdateli/El_izd/2021-34.pdf

This students-friendly tutorial presents the fundamental principles of physics in a clear and concise manner. Emphasizing conceptual understanding as the basis for mastering a variety of problem-solving tools, it provides a wide range of relevant applications and illustrative examples. This book discusses mechanics, thermodynamics, oscillations and wave motion, electricity and magnetism, optics, and atomic physics.

For the students learning more about the fundamentals of physics and how physical principles apply to a variety of real-world situations, devices.

При подготовке электронного издания использовались следующие программные средства:

- Adobe Acrobat – текстовый редактор;
- Microsoft Word – текстовый редактор.

Авторы: Alexandrov Igor V., Vasil’ev Denis Y.

Редактирование и верстка: *Ф. Х. Гарипова*
Программирование и компьютерный дизайн: *А. П. Меркулова*

ФГБОУ ВО «Уфимский государственный авиационный технический университет»
450008, Уфа, ул. К. Маркса, 12.

Тел.: +7-908-35-05-007
e-mail: rik.ugatu@yandex.ru

INTRODUCTION

Science consists of the theories and laws that are the general truths of nature. *Physics* is concerned with describing the interactions of energy, matter, space, and time, and it is especially interested in what fundamental mechanisms underlie every phenomenon. The concern for describing the basic phenomena in nature essentially defines the *realm of physics*. Physics aims to describe the function of everything around us, from the movement of tiny charged particles to the motion of people, cars, and spaceships. In fact, almost everything around you can be described quite accurately by the laws of physics.

Physics is the foundation of many important disciplines and contributes directly to others. Chemistry, for example – since it deals with the interactions of atoms and molecules – is rooted in atomic and molecular physics. Most branches of engineering are applied physics. In architecture, physics is at the heart of structural stability, and is involved in the acoustics, heating, lighting, and cooling of buildings. Parts of geology rely heavily on physics, such as radioactive dating of rocks, earthquake analysis, and heat transfer in the Earth. Some disciplines, such as biophysics and geophysics, are hybrids of physics and other disciplines.

The *laws of nature* are concise descriptions of the universe around us; they are human statements of the underlying laws or rules that all natural processes follow. Such laws are intrinsic to the universe; humans did not create them and so cannot change them. We can only discover and understand them. Their discovery is a very human endeavor, with all the elements of mystery, imagination, struggle, triumph, and disappointment inherent in any creative effort. The cornerstone of discovering natural laws is observation; science must describe the universe as it is, not as we may imagine it to be.

The range of objects and phenomena studied in physics is immense. From the incredibly short lifetime of a nucleus to the age of the Earth, from the tiny sizes of sub-nuclear particles to the vast distance to the edges of the known universe, from the force exerted by a jumping flea to the force between Earth and the Sun. Giving numerical values for physical quantities and equations for physical principles allows us to understand nature much more deeply than does qualitative description alone. To comprehend these vast ranges, we must also have accepted units in which to express them. And we shall find that (even in the potentially mundane discussion of meters, kilograms, and seconds) a profound simplicity of nature appears – all physical quantities can be expressed as combinations of only four fundamental physical quantities: length, mass, time, and electric current. We define a physical quantity either by specifying how it is measured or by stating how it is calculated from other measurements. For example, we define distance and time by specifying methods for measuring them, whereas we define average speed by stating that it is calculated as distance traveled divided by time of travel. Measurements of physical quantities are expressed in terms of units, which are standardized values.

1. THEORY

1.1. Mechanics

The study of physics begins with *kinematics* which is defined as the study of motion without considering its causes. The word “kinematics” comes from a Greek term meaning motion. In one-dimensional kinematics and two-dimensional kinematics we will study only the motion of a football, for example, without worrying about what forces cause or change its motion. In order to describe the motion of an object, you must first be able to describe its *position* – where it is at any particular time. More precisely, you need to specify its position relative to a convenient reference frame. Earth is often used as a reference frame, and we often describe the position of an object as it relates to stationary objects in that reference frame. For example, a rocket launch would be described in terms of the position of the rocket with respect to the Earth. In other cases, we use reference frames that are not stationary but are in motion relative to the Earth. To describe the position of a person in an airplane, for example, we use the airplane, not the Earth, as the reference frame. If an object moves relative to a reference frame, then the object’s position changes. This change in position is known as displacement. The word “displacement” implies that an object has moved, or has been displaced. *Displacement* is the change in position of an object:

$$\Delta x = x_f - x_0, \tag{1.1.1}$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

In physics, the definition of time is simple – *time* is change, or the interval over which change occurs. It is impossible to know that time has passed unless something changes. The amount of time or change is calibrated by comparison with a standard. The SI (this acronym is derived from the French *Système International*) unit for time is the second, abbreviated [s]. *Elapsed time* Δt is the difference between the ending time and beginning time:

$$\Delta t = t_f - t_0, \tag{1.1.2}$$

where Δt is the change in time or elapsed time, t_f is the time at the end of the motion, and t_0 is the time at the beginning of the motion (Δ means the change in the quantity that follows it).

In everyday language, most people use the terms “speed” and “velocity” interchangeably. In physics, however, they do not have the same meaning and they are distinct concepts. One major difference is that speed has no direction. Thus speed is a scalar. Just as we need to distinguish between instantaneous velocity and average velocity, we also need to distinguish between instantaneous speed and average speed. *Instantaneous speed* is the magnitude of instantaneous velocity. *Average speed* $\bar{v}$ is the distance traveled divided by elapsed time:

$$\bar{v} = \frac{\Delta x}{\Delta t}. \tag{1.1.3}$$

Another important value in mechanics – *average acceleration* $\bar{a}$ is the rate at which velocity changes:

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v_f - v_0}{t_f - t_0}, \tag{1.1.4}$$

where v is velocity, and t is time.

Solving for Displacement Δx and Final Position x from Average Velocity when Acceleration a is Constant:

$$\bar{v} = \frac{x - x_0}{t}, \tag{1.1.5}$$

$$x = x_0 + \bar{v}t, \tag{1.1.6}$$

$$\bar{v} = \frac{v_0 + v}{2} (\text{const}\, a). \tag{1.1.7}$$

Solving for final velocity:

$$a = \frac{v - v_0}{t} (\text{const}\, a), \tag{1.1.8}$$

$$v = v_0 a t (\text{const}\, a). \tag{1.1.9}$$

Solving for final position, when velocity is not constant ($a \neq 0$):

$$\bar{v} = v_0 + \frac{1}{2}at, \tag{1.1.10}$$

$$x = x_0 + v_0t + \frac{1}{2}at^2(\text{const } a). \tag{1.1.11}$$

Falling objects form an interesting class of motion problems. The most remarkable and unexpected fact about falling objects is that, if air resistance and friction are negligible, then in a given location all objects fall toward the center of Earth with the same constant acceleration, independent of their mass. Kinematic equations for objects in free-fall, where acceleration $a = -g$ (≈9.81 m/s^2):

$$v = v_0 - gt, \tag{1.1.12}$$

$$y = y_0 + v_0t - \frac{1}{2}gt^2, \tag{1.1.13}$$

$$v^2 = v_0^2 - 2g(y - y_0). \tag{1.1.14}$$

This equation defines the maximum height h of a projectile and depends only on the vertical component y of the initial velocity:

$$h = \frac{v_{0y}^2}{2g}. \tag{1.1.15}$$

The study of motion is kinematics, but kinematics only describes the way objects move – their velocity and their acceleration. *Dynamics* considers the forces that affect the motion of moving objects and systems. Newton's laws of motion are the foundation of dynamics. These laws provide an example of the breadth and simplicity of principles under which nature functions. They are also universal laws in that they apply to similar situations on Earth as well as in space. It was not until the advent of modern physics early in the 20th century that it was discovered that Newton's laws of motion produce a good approximation to motion only when the objects are moving at speeds much, much less than the speed of light and when those objects are larger than the size of most molecules ($\approx 10^{-9}$ m in diameter). These constraints define the realm of classical mechanics. *Dynamics* is the study of the forces that cause objects and systems to move. To understand this, we need a working definition of force. Our intuitive definition of *force* – that is, a push or a pull – is a good place to start. We know that a push or pull has both magnitude and direction (therefore, it is a vector quantity) and can vary considerably in each regard.

Experience suggests that an object at rest will remain at rest if left alone, and that an object in motion tends to slow down and stop unless some effort is made to keep it moving. What *Newton's first law of motion* states, however, is the following: a body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force. The property of a body to remain at rest or to remain in motion with constant velocity is called *inertia*. Newton's first law is often called *the law of inertia*. As we know from experience, some objects have more inertia than others. The inertia of an object is measured by its mass m. Roughly speaking, *mass* is a measure of the amount of "stuff" (or matter) in something. The quantity or amount of matter in an object is determined by the numbers of atoms and molecules of various types it contains. Unlike weight, mass does not vary with location. The mass of an object is the same on Earth, in orbit, or on the surface of the Moon. In practice, it is very difficult to count and identify all of the atoms and molecules in an object, so masses are not often determined in this manner. Operationally, the masses of objects are determined by comparison with the standard kilogram.

Newton's second law of motion is closely related to Newton's first law of motion. It mathematically states the cause and effect relationship between force F and changes in motion. Newton's second law of motion is more quantitative and is used extensively to calculate what happens in situations involving a force: The acceleration of a system is directly proportional to and in the same direction as the net external force acting on the system, and inversely proportional to its mass. In equation form, Newton's second law of motion is:

$$\bar{a} = \frac{\bar{F}}{m}. \tag{1.1.16}$$

This is the equation for *weight* – the gravitational force on a mass m:

$$\bar{w} = m\bar{g}. \tag{1.1.17}$$

Recall that g can take a positive or negative value, depending on the positive direction in the coordinate system. Mass and weight are often used interchangeably in everyday language. However, in science, these terms are distinctly different from one another. Mass is a measure of how much matter is in an object. The typical measure of mass is the kilogram (or the “slug” in English units). Weight, on the other hand, is a measure of the force of gravity acting on an object. Weight is equal to the mass of an object [m] multiplied by the acceleration due to gravity [g]. Like any other force, weight is measured in terms of newtons [N = kg·m/s^2] (or pounds in English units).

Newton’s third law of motion is the symmetry in forces: whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts. In equation form:

$$\bar{F}_{1,2} = -\bar{F}_{2,1}. \tag{1.1.18}$$

Forces are given many names, such as push, pull, thrust, lift, weight, friction, and tension. Traditionally, forces have been grouped into several categories and given names relating to their source, how they are transmitted, or their effects. *Weight* (also called force of gravity) is a pervasive force that acts at all times and must be counteracted to keep an object from falling. A *tension* is a force along the length of a medium, especially a force carried by a flexible medium, such as a rope or cable. The word “tension” comes from a Latin word meaning “to stretch”. Not coincidentally, the flexible cords that carry muscle forces to other parts of the body are called *tendons*. Any flexible connector, such as a string, rope, chain, wire, or cable, can exert pulls only parallel to its length; thus, a force carried by a flexible connector is a tension with direction parallel to the connector. It is important to understand that tension is a pull in a connector.

There is another distinction among forces in addition to the types already mentioned. Some forces are real, whereas others are not. *Real forces* are those that have some physical origin, such as the gravitational pull. Contrastingly, *fictitious forces* are those that arise simply because an observer is in an accelerating frame of reference, such as one that rotates (like a merry-go-round) or undergoes linear acceleration (like a car slowing down). For example, if a satellite is heading due north above Earth's northern hemisphere, then to an observer on Earth it will appear to experience a force to the west that has no physical origin. Of course, what is happening here is that Earth is rotating toward the east and moves east under the satellite. In Earth's frame this looks like a westward force on the satellite, or it can be interpreted as a violation of Newton's first law (the law of inertia). An *inertial frame of reference* is one in which all forces are real and, equivalently, one in which Newton's laws have the simple forms. Earth's rotation is slow enough that Earth is nearly an inertial frame. You ordinarily must perform precise experiments to observe fictitious forces and the slight departures from Newton's laws, such as the effect just described. On the large scale, such as for the rotation of weather systems and ocean currents, the effects can be easily observed. The crucial factor in determining whether a frame of reference is inertial is whether it accelerates or rotates relative to a known inertial frame. Unless stated otherwise, all phenomena discussed in this text are considered in inertial frames.

One of the most remarkable simplifications in physics is that only four distinct forces account for all known phenomena. In fact, nearly all of the forces we experience directly are due to only one basic force, called the electromagnetic force (the gravitational force is the only force we experience directly that is not electromagnetic). This is a tremendous simplification of the myriad of apparently different forces we can list, only a few of which were discussed in the previous section. As we will see, the basic forces are all thought to act through the exchange of microscopic carrier particles, and the characteristics of the basic forces are determined by the types of particles exchanged. Action at a distance, such as the gravitational force of Earth on the Moon, is explained by the existence of a *force field* rather than by "physical contact". The *four basic forces* are the gravitational force, the electromagnetic force, the weak nuclear force, and the strong nuclear force. Their properties are summarized in Table 1. Since the weak and strong nuclear forces act over an extremely short range, the size of a nucleus or less, we do not experience them directly, although they are crucial to the very structure of matter. These forces determine which nuclei are stable and which decay, and they are the basis of the release of energy in certain nuclear reactions. Nuclear forces determine not only the stability of nuclei, but also the relative abundance of elements in nature.

Table 1

Properties of the Four Basic Forces

Force	Approximate Relative Strengths	Range	Attraction/Repulsion	Carrier Particle
Gravitational	10^{-38}	∞	attractive only	gravitation
Electromagnetic	10^{-2}	∞	attractive and repulsive	photon
Weak nuclear	10^{-13}	$< 10^{-18}$ m	attractive and repulsive	W^+, W^-, Z^0
Strong nuclear	1	$< 10^{-15}$ m	attractive and repulsive	gluons

Attempts to unify the four basic forces are discussed in relation to elementary particles. By "unify" we mean finding connections between the forces that show that they are different manifestations of a single force. Even if such unification is achieved, the forces will retain their separate characteristics on the macroscopic scale and may be identical only under extreme conditions such as those existing in the early universe. The concept of a *force field* is also used in connection with electric charge. It is also a useful idea for all the basic forces. Fields help us to visualize forces and how they are transmitted, as well as to describe them with precision and to link forces with subatomic carrier particles. The *graviton* is a proposed particle, though it has not yet been observed by scientists. The particles W^+, W^-, and Z^0 are called *vector bosons*. These were predicted by theory and first observed in 1983. There are eight types of *gluons* proposed by scientists, and their existence is indicated by meson exchange in the nuclei of atoms.

Friction is a force that is around us all the time that opposes relative motion between systems in contact but also allows us to move (which you have discovered if you have ever tried to walk on ice). While a common force, the behavior of friction is actually very complicated and is still not completely understood. *Friction* is a force that opposes relative motion between systems in contact. Magnitude of static friction *fs* is:

$$f_s \leq \mu_s N, \tag{1.1.19}$$

where μ_s is the coefficient of static friction and N is the magnitude of the normal force.

Another interesting force in everyday life is the force of drag on an object when it is moving in a fluid (either a gas or a liquid). Like friction, the *drag force* always opposes the motion of an object. Unlike simple friction, the drag force is proportional to some function of the velocity of the object in that fluid. This functionality is complicated and depends upon the shape of the object, its size, its velocity, and the fluid it is in. Drag force F_D is found to be proportional to the square of the speed of the object. Mathematically:

$$F_D \sim v^2 \tag{1.1.20}$$

or

$$F_D = \frac{1}{2} C \rho A v^2, \tag{1.1.21}$$

where C is the drag coefficient, A is the area of the object facing the fluid, and ρ is the density of the fluid.

We now move from consideration of forces that affect the motion of an object (such as friction and drag) to those that affect an object's shape. If a bulldozer pushes a car into a wall, the car will not move but it will noticeably change shape. A change in shape due to the application of a force is a *deformation*. Even very small forces are known to cause some deformation. For small deformations, two important characteristics are observed. First, the object returns to its original shape when the force is removed – that is, the deformation is elastic for small deformations. Second, the size of the deformation is proportional to the force – that is, for small deformations, Hooke's law is obeyed. In equation form, *Hooke's law* is given by:

$$F = k \Delta L, \tag{1.1.22}$$

where ΔL is the amount of deformation (the change in length, for example) produced by the force F, and k is a proportionality constant that depends on the shape and composition of the object and the direction of the force. Note that this force is a function of the deformation ΔL – it is not constant as a kinetic friction force is.

In Kinematics, we studied motion along a straight line and introduced such concepts as displacement, velocity, and acceleration. Two-Dimensional Kinematics dealt with motion in two dimensions.

Projectile motion is a special case of two-dimensional kinematics in which the object is projected into the air, while being subject to the gravitational force, and lands a distance away. Now, we consider situations where the object does not land but moves in a curve. We begin the study of uniform circular motion by defining two angular quantities needed to describe rotational motion. The *arc length Δs* is the distance traveled along a circular path as shown in Fig. 1.

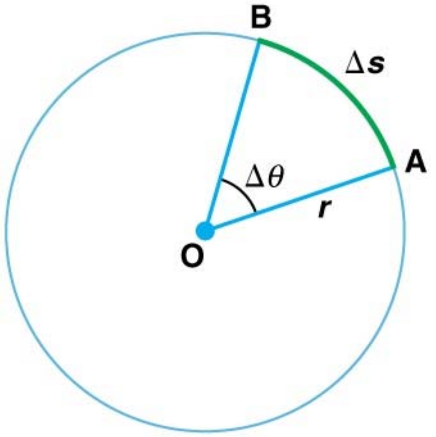

Fig. 1. The radius of a circle is rotated through an angle *Δ*θ.
The arc length Δ*s* is described on the circumference

Note that *r* is the *radius of curvature* of the circular path. We know that for one complete revolution, the arc length is the circumference of a circle of radius *r*. The circumference of a circle is 2*πr*. Thus for one complete revolution the rotation angle is:

$$\Delta\theta = \frac{2\pi r}{r} = 2\pi. \tag{1.1.23}$$

This result is the basis for defining the units used to measure rotation angles, Δθ to be radians [rad], defined so that:

$$2\pi\, rad = 1 \text{ revolurion.}$$

How fast is an object rotating? We define *angular velocity ω* as the rate of change of an angle. In symbols, this is:

$$\omega = \frac{\Delta\theta}{\Delta t}, \tag{1.1.24}$$

where an angular rotation $\Delta\theta$ takes place in a time Δt. The greater the rotation angle in a given amount of time, the greater the angular velocity. The units for angular velocity are radians per second [rad/s].

From $\Delta\theta = \frac{\Delta s}{r}$ we see that $\Delta s = r\Delta\theta$. Substituting this into the expression for v gives:

$$v = \frac{r\Delta\theta}{\Delta t} = r\omega. \qquad (1.1.25)$$

In uniform circular motion, the direction of the velocity changes constantly, so there is always an associated acceleration, even though the magnitude of the velocity might be constant. We call the acceleration of an object moving in uniform circular motion (resulting from a net external force) the *centripetal acceleration* a_c; centripetal means "toward the center" or "center seeking". Finally, noting that $\frac{\Delta v}{\Delta t} = a_c$ and that $\frac{\Delta s}{\Delta t} = v$,, the linear or tangential speed, we see that the magnitude of the centripetal acceleration is:

$$a_c = \frac{v^2}{r} = r\omega^2, \qquad (1.1.26)$$

which is the acceleration of an object in a circle of radius r at a speed v.

Any force or combination of forces can cause a centripetal or radial acceleration. Just a few examples are the tension in the rope on a tether ball, the force of Earth's gravity on the Moon, friction between roller skates and a rink floor, a banked roadway's force on a car, and forces on the tube of a spinning centrifuge. Any net force causing uniform circular motion is called a *centripetal force*. The direction of a centripetal force is toward the center of curvature, the same as the direction of centripetal acceleration. For uniform circular motion, the acceleration is the centripetal acceleration – $a = a_c$. Thus, the magnitude of centripetal force F_c is:

$$F_c = ma_c = m\frac{v^2}{r} = mr\omega^2. \qquad (1.1.27)$$

The gravitational force is relatively simple. It is always attractive, and it depends only on the masses involved and the distance between them. Stated in modern language, *Newton's universal law of gravitation* states that every particle in the universe attracts every other particle with a force along a line joining them. The force is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. The bodies we are dealing with tend to be large. To simplify the situation we assume that the body acts as if its entire mass is concentrated at one specific point called the *center of mass*. For two bodies having masses m and M with a distance r between their centers of mass, the equation for Newton's universal law of gravitation is:

$$F = G\frac{mM}{r^2}, \tag{1.1.28}$$

where F is the magnitude of the gravitational force and G is a proportionality factor called the *gravitational constant*. G is a universal gravitational constant – that is, it is thought to be the same everywhere in the universe. It has been measured experimentally to be:

$$G = 6.673\cdot 10^{-11}\ [N\cdot \mathrm{m}^2/\mathrm{kg}^2].$$

Energy plays an essential role both in everyday events and in scientific phenomena. Not only does energy have many interesting forms, it is involved in almost all phenomena, and is one of the most important concepts of physics. What makes it even more important is that the total amount of energy in the universe is constant. Energy can change forms, but it cannot appear from nothing or disappear without a trace. Energy is thus one of a handful of physical quantities that we say is conserved. *Conservation of energy* (as physicists like to call the principle that energy can neither be created nor destroyed) is based on experiment. Even as scientists discovered new forms of energy, conservation of energy has always been found to apply. There is no simple, yet accurate, scientific definition for energy. Energy is characterized by its many forms and the fact that it is conserved. We can loosely define *energy* as the ability to do work, admitting that in some circumstances not all energy is available to do work. Because of the association of energy with work, we begin the chapter with a discussion of work. Work is intimately related to energy and how energy moves from one system to another or changes form. For *work*, in the scientific sense, to be done, a force must be exerted and there must be motion or displacement in the direction of the force. Formally, the *work* done on a system by a constant force is defined to be the product of the component of the force in the direction of motion times the distance through which the force acts. For one-way motion in one dimension, this is expressed in equation form as:

$$W = Fd\cos\theta, \tag{1.1.29}$$

where W is work, d is the displacement of the system, and θ is the angle between the force vector F and the displacement vector d.

Work and energy have the same units. From the definition of work, we see that those units are force times distance. Thus, in SI units, work and energy are measured in *newton-meters*. A newton-meter is given the special name *joule* [$J = N{\cdot}m = \mathrm{kg{\cdot}m^2/s^2}$]. One joule is not a large amount of energy; it would lift a small 100-gram apple a distance of about 1 meter.

We know from the study of Newton’s laws in Dynamics: Force and Newton's Laws of Motion that net force causes acceleration. Work done by the net force gives, system energy of motion, and in the process we will also find an expression for the energy of motion. Network is defined to be the sum of work done by all external forces – that is, network is the work done by the net external force F_{net} . In equation form, this is:

$$W_{net} = F_{net}d\cos\theta, \tag{1.1.30}$$

where θ is the angle between the force vector and the displacement vector. The network on a system equals the change in the quantity $\frac{1}{2}mv^2$:

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2. \tag{1.1.31}$$

The quantity $\frac{1}{2}mv^2$ in the work-energy theorem is defined to be the translational *kinetic energy* (*KE*) of a mass m moving at a speed v. In equation form, the translational kinetic energy:

$$KE = \frac{1}{2}mv^2, \tag{1.1.32}$$

is the energy associated with translational motion. Kinetic energy is a form of energy associated with the motion of a particle, single body, or system of objects moving together.

Lifting objects is work in both the scientific and everyday sense – it is work done against the gravitational force. When there is work, there is a transformation of energy. The work done against the gravitational force goes into an important form of stored energy. If the object is lifted straight up at constant speed, then the force needed to lift it is equal to its weight mg. The work done on the mass is then

$$W = Fd = mgh. \tag{1.1.33}$$

We define this to be the *gravitational potential energy* (*PEg*) put into or gained by the object-Earth system. This energy is associated with the state of separation between two objects that attract each other by the gravitational force. For convenience, we refer to this as the *PEg* gained by the object, recognizing that this is energy stored in the gravitational field of Earth. Gravitational potential energy may be converted to other forms of energy, such as kinetic energy. Work is done by a force, and some forces, such as weight, have special characteristics. A *conservative force* is one, like the gravitational force, for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken. We can define a *potential energy* (*PE*) for any conservative force, just as we did for the gravitational force. Potential energy is the energy a system has due to position, shape, or configuration. It is stored energy that is completely recoverable. A *conservative force* is one for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken. We can define a *potential energy* (*PE*) for any conservative force. The work done against a conservative force to reach a final configuration depends on the configuration, not the path followed, and is the potential energy added.

First, let us obtain an expression for the *potential energy stored* in a spring (*PEs*). We calculate the work done to stretch or compress a spring than obeys Hooke's law. For our spring the amount of deformation produced by a force F by the distance x that the spring is stretched or compressed along its length. So the force needed to stretch the spring has magnitude $F = kx$, where k is the *spring's force constant*. The force increases linearly from 0 at the start to kx in the fully stretched position.

The average force is $kx/2$, thus the work done in stretching or compressing the spring is:

$$W_s = Fd = \left(\frac{kx}{2}\right)x = \frac{1}{2}kx^2. \qquad (1.1.34)$$

We therefore define the potential energy of a spring, PE_s, to be

$$PE_s = \frac{1}{2}kx^2, \qquad (1.1.35)$$

where k is the spring's force constant and x is the displacement from its undeformed position. The potential energy represents the work done on the spring and the energy stored in it as a result of stretching or compressing it a distance x. The potential energy of the spring PE_s does not depend on the path taken, it depends only on the stretch or squeeze x in the final configuration.

Let us now consider what form the work-energy theorem takes when only conservative forces are involved. This will lead us to the conservation of energy principle. The work-energy theorem states that the net work done by all forces acting on a system equals its change in kinetic energy. In equation form, this is:

$$W_{net} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 = \Delta KE. \qquad (1.1.36)$$

If only conservative forces act, then

$$W_{net} = W_c \qquad (1.1.37)$$

where W_c is the total work done by all conservative forces. Thus,

$$W_c = \Delta KE. \qquad (1.1.38)$$

Now, if the conservative force, such as the gravitational force or a spring force, does work, the system loses potential energy. That is, $W_c = -\Delta PE$. Therefore,

$$-\Delta PE = \Delta KE, \tag{1.1.39}$$

or

$$\Delta KE + \Delta PE = 0. \tag{1.1.40}$$

This equation means that the total kinetic and potential energy is constant for any process involving only conservative forces. That is:

$$\left.\begin{array}{c} KE + PE = \text{constant} \\ \text{or} \\ KE_i + PE_i = KE_f + PE_f \end{array}\right\} \text{conservate force only,} \tag{1.1.41}$$

where i and *f* denote initial and final values. This equation is a form of the work-energy theorem for conservative forces, it is known as the *conservation of mechanical energy* principle. Remember that this applies to the extent that all the forces are conservative, so that friction is negligible. The total kinetic plus potential energy of a system is defined to be its *mechanical energy* (*KE* + *PE*). In a system that experiences only conservative forces, there is a potential energy associated with each force, and the energy only changes form between *KE* and the various types of *PE*, with the total energy remaining constant.

Forces are either conservative or nonconservative. A *nonconservative force* is one for which work depends on the path taken. *Friction* is a good example of a nonconservative force. An important characteristic is that the work done by a nonconservative force adds or removes mechanical energy from a system. Friction, for example, creates *thermal energy* that dissipates, removing energy from the system. Furthermore, even if the thermal energy is retained or captured, it cannot be fully converted back to work, so it is lost or not recoverable in that sense as well.

Friction stops the player by converting his kinetic energy into other forms, including thermal energy. In terms of the work-energy theorem, the work done by friction, which is negative, is added to the initial kinetic energy to reduce it to zero. The work done by friction is negative, because *f* is in the opposite direction of the motion (that is, $\theta = \pi$, and so $\cos\theta = -1$). Thus $W_{nc} = -fd$. The equation simplifies to

$$\frac{1}{2}mv_i^2 - fd = 0. \tag{1.1.42}$$

The fact that energy is conserved and has many forms makes it very important. What are some other forms of energy? Let us detail a few here. *Electrical energy* is a common form that is converted to many other forms and does work in a wide range of practical situations. Fuels, such as gasoline and food, carry *chemical energy* that can be transferred to a system through oxidation. Chemical fuel can also produce electrical energy, such as in batteries. Batteries can in turn produce light, which is a very pure form of energy. Most energy sources on Earth are in fact stored energy from the energy we receive from the Sun. We sometimes refer to this as *radiant energy*, or electromagnetic radiation, which includes visible light, infrared, and ultraviolet radiation. *Nuclear energy* comes from processes that convert measurable amounts of mass into energy. Nuclear energy is transformed into the energy of sunlight, into electrical energy in power plants, and into the energy of the heat transfer and blast in weapons. Atoms and molecules inside all objects are in random motion. This internal mechanical energy from the random motions is called *thermal energy*, because it is related to the temperature of the object. These and all other forms of energy can be converted into one another and can do work. The transformation of energy from one form into others is happening all the time. The chemical energy in food is converted into thermal energy through metabolism; light energy is converted into chemical energy through photosynthesis. In a larger example, the chemical energy contained in coal is converted into thermal energy as it burns to turn water into steam in a boiler. This thermal energy in the steam in turn is converted to mechanical energy as it spins a turbine, which is connected to a generator to produce electrical energy. In all of these examples, not all of the initial energy is converted into the forms mentioned. Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell produces electricity, which in turn can be used to run an electric motor. Energy is converted from the primary source of solar energy into electrical energy and then into mechanical energy.

Even though energy is conserved in an energy conversion process, the output of useful energy or work will be less than the energy input. The *efficiency* E_{ff} of an energy conversion process is defined as

$$E_{ff} = \frac{\text{useful energy or work output}}{\text{total energy input}} = \frac{W_{out}}{E_{in}}. \tag{1.1.43}$$

The Table 2 gives the amount of energy stored, used, or released from various objects and in various phenomena.

Table 2

Energy of Various Objects and Phenomena

Object/Phenomenon	Energy in Joules
Big Bang	10^{68}
Energy released in a supernova	10^{44}
Fusion of all the hydrogen in Earth’s oceans	10^{34}
Annual world energy use	$4 \cdot 10^{20}$
Large fusion bomb (9 megaton)	$3.8 \cdot 10^{16}$
One kg uranium (nuclear fission)	$8 \cdot 10^{13}$
One gallon of gasoline	$1.2 \cdot 10^{8}$
Daily home electricity use (developed countries)	$7 \cdot 10^{7}$
One g protein (4.1 kcal)	$1.7 \cdot 10^{4}$
Tennis ball at 100 km/h	22
Single electron in a TV tube beam	$4 \cdot 10^{-15}$
Energy to break one DNA strand	10^{-19}

Table 3 lists some efficiencies of mechanical devices and human activities. The other 60 % transforms into other (perhaps less useful) energy forms, such as thermal energy, which is then released to the environment through combustion gases and cooling towers.

Table 3

Efficiency of the Human Body and Mechanical Devices

Activity/Device	Efficiently in %
Electric motor	98
Coal power plant	42
Shoveling	3

Power P – is the rate at which work is done:

$$P = \frac{W}{t}. \tag{1.1.44}$$

The SI unit for power is the watt [W], where 1 watt equals 1 joule/second [1 W = 1 J/s]. Because work is energy transfer, power is also the rate at which energy is expended. A 60-W light bulb, for example, expends 60 J of energy per second. Great power means a large amount of work or energy developed in a short time. For example, when a powerful car accelerates rapidly, it does a large amount of work and consumes a large amount of fuel in a short time. Examples of power are limited only by the imagination, because there are as many types as there are forms of work and energy. See Table 4 for some examples.

Table 4

Power Output or Consumption

Object or Phenomenon	Power in Watts
Supernova at peak	$5 \cdot 10^{37}$
Milky Way galaxy	10^{37}
The Sun	$4 \cdot 10^{26}$
Lightning bolt	$2 \cdot 10^{12}$
Clothes dryer	$4 \cdot 10^{3}$
Heart, person at rest (total useful and heat transfer)	8
Pocket calculator	10^{-3}

The scientific definition of linear momentum is consistent with most people's intuitive understanding of momentum: a large, fast-moving object has greater momentum than a smaller, slower object. *Linear momentum p* is defined as the product of a system's mass multiplied by its velocity. In symbols, linear momentum is expressed as

$$\vec{p} = m\vec{v}, \tag{1.1.45}$$

Momentum is directly proportional to the object's mass and also its velocity. Thus the greater an object's mass or the greater its velocity, the greater its momentum. Momentum p is a vector having the same direction as the velocity v. The SI unit for momentum is [kg · m/s]. The importance of momentum, unlike the importance of energy, was recognized early in the development of classical physics. Momentum was deemed so important that it was called the "*quantity of motion*". Newton actually stated his second law of motion in terms of momentum: The net external force equals the change in momentum of a system divided by the time over which it changes. Using symbols, this law is

$$\overrightarrow{F_{net}} = \frac{d\vec{p}}{dt}, \tag{1.1.46}$$

where F_{net} is the net external force, dp is the change in momentum, and dt is the change in time.

The effect of a force on an object depends on how long it acts, as well as how great the force is. Quantitatively, the effect we are talking about is the change in momentum dp. Total momentum is conserved for any isolated system, with any number of objects in it. In equation form, the conservation of momentum principle for an isolated system is written

$$p_{tot} = \text{constant}, \tag{1.1.47}$$

where p_{tot} is the total momentum, the sum of the momenta of the individual objects in the system.

An *inelastic collision* is one in which the internal kinetic energy changes (it is not conserved). This lack of conservation means that the forces between colliding objects may remove or add internal kinetic energy. Work done by internal forces may change the forms of energy within a system. For inelastic collisions, such as when colliding objects stick together, this internal work may transform some internal kinetic energy into heat transfer. Or it may convert stored energy into internal kinetic energy, such as when exploding bolts separate a satellite from its launch vehicle.

Statics is the study of forces in equilibrium, a large group of situations that makes up a special case of Newton's second law. The first condition necessary to achieve equilibrium is the one already mentioned: the net external force on the system must be zero. Expressed as an equation, this is simply

$$F_{net} = 0. \tag{1.1.48}$$

Note that if net F is zero, then the net external force in any direction is zero. The second condition necessary to achieve equilibrium involves avoiding accelerated rotation (maintaining a constant angular velocity. A rotating body or system can be in equilibrium if its rate of rotation is constant and remains unchanged by the forces acting on it. The magnitude, direction, and point of application of the force are incorporated into the definition of the physical quantity called torque. *Torque* is the rotational equivalent of a force. It is a measure of the effectiveness of a force in changing or accelerating a rotation (changing the angular velocity over a period of time). In equation form, the magnitude of torque is defined to be

$$M = rF\sin\theta, \tag{1.1.49}$$

where M is the symbol for torque, r is the distance from the pivot point to the point where the force is applied, F is the magnitude of the force, and θ is the angle between the force and the vector directed from the point of application to the pivot point, as seen in Fig. 2.

Fig. 2. Measurement of Torque

Before we can consider the rotation of anything other than a point mass, we must extend the idea of rotational inertia to all types of objects. To expand our concept of rotational inertia, we define the *moment of inertia* I of an object to be the sum of mr^2 for all the point masses of which it is composed. That is, $I = mr^2$. Here I is analogous to m in translational motion. Note that I has units of mass multiplied by distance squared [$\mathrm{kg} \cdot \mathrm{m}^2$], as we might expect from its definition. The general relationship among torque, moment of inertia, and angular acceleration is

$$\vec{M} = I\vec{\varepsilon}. \tag{1.1.50}$$

where M is the total torque from all forces relative to a chosen axis and ε is angular acceleration. For simplicity, we will only consider torques exerted by forces in the plane of the rotation. Such torques are either positive or negative and add like ordinary numbers. The relationship in (1.1.50) I is the rotational analog to Newton's second law and is very generally applicable. This equation is actually valid for any torque, applied to any object, relative to any axis.

Work and energy in rotational motion are completely analogous to work and energy in translational motion, first presented in *Uniform Circular Motion and Gravitation*. Equation (1.1.51) is the *work-energy theorem* for rotational motion only. As you may recall, network changes the kinetic energy of a system. Through an analogy with translational motion, we define the term $\frac{1}{2}I\omega^2$ to be *rotational kinetic energy* KE_{rot} for an object with a moment of inertia I and an angular velocity ω:

$$KE_{rot} = \frac{1}{2}I\omega^2. \tag{1.1.51}$$

The expression for rotational kinetic energy is exactly analogous to translational kinetic energy, with *I* being analogous to m and ω to *v*.

By now the pattern is clear – every rotational phenomenon has a direct translational analog. It seems quite reasonable, then, to define *angular momentum L* as

$$\vec{L} = [\vec{r}, \vec{p}] \quad \text{or} \quad \vec{L} = I\vec{\omega}. \tag{1.1.52}$$

This equation is an analog to the definition of linear momentum as *p*. Units for linear momentum are [kg · m/s] while units for angular momentum are [kg · m^2/s]. As we would expect, an object that has a large moment of inertia *I*, such as Earth, has a very large angular momentum. An object that has a large angular velocity ω, such as a centrifuge, also has a rather large angular momentum.

1.2. Molecular Physics and Thermodynamics

Heat is something familiar to each of us. We feel the warmth of the summer Sun, the chill of a clear summer night, the heat of coffee after a winter stroll, and the cooling effect of our sweat. Heat transfer is maintained by temperature differences. Manifestations of heat transfer – the movement of heat energy from one place or material to another – are apparent throughout the universe. Heat from beneath Earth's surface is brought to the surface in flows of incandescent lava. The Sun warms Earth's surface and is the source of much of the energy we find on it. Rising levels of atmospheric carbon dioxide threaten to trap more of the Sun's energy, perhaps fundamentally altering the ecosphere. In space, supernovas explode, briefly radiating more heat than an entire galaxy does.

The concept of temperature has evolved from the common concepts of hot and cold. The scientific definition of temperature is less ambiguous than your senses of hot and cold. *Temperature* is operationally defined to be what we measure with a *thermometer*. Many physical quantities are defined solely in terms of how they are measured. Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer.

Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip other properties used to measure temperature include electrical resistance and color. Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the *Fahrenheit*, *Celsius*, and *Kelvin scales*. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used. The Celsius scale (which replaced the slightly different centigrade scale) has the freezing point of water at 0 ºC and the boiling point at 100 ºC. Its unit is the degree Celsius (ºC). On the Fahrenheit scale (still the most frequently used in the United States), the freezing point of water is at 32 ºF and the boiling point is at 212 ºF. The unit of temperature on this scale is the degree Fahrenheit (ºF). Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale 180/100 = 9/5. The Kelvin scale is the temperature scale that is commonly used in science. It is an absolute temperature scale defined to have 0 K at the lowest possible temperature, called absolute zero. The official temperature unit on this scale is the kelvin, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work. The relationship between the three common temperature scales is shown in Fig. 3. Temperatures on these scales can be converted using the equations in Table 5.

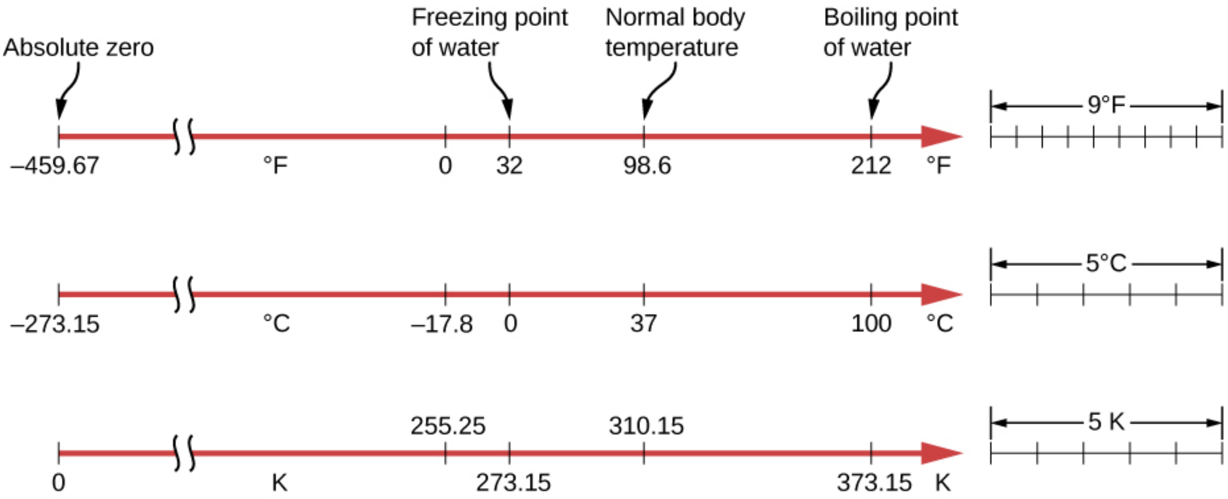

Fig. 3. Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree

The lowest temperatures ever recorded have been measured during laboratory experiments: $4.5 \cdot 10^{-10}$ K at the Massachusetts Institute of Technology (USA), and $1.0 \cdot 10^{-10}$ K at Helsinki University of Technology (Finland). In comparison, the coldest recorded place on Earth's surface is Vostok (Russian scientific station) in Antarctica at 183 K (–89 ºC), and the coldest place (outside the lab) known in the universe is the Boomerang Nebula, with a temperature of 1 K.

Table 5

Temperature Conversions

To convert from	Use this equation
Celsius to Fahrenheit	$T(°\mathrm{F}) = \frac{9}{5}T(°\mathrm{C}) + 32$
Fahrenheit to Celsius	$T(°\mathrm{C}) = \frac{5}{9}(T(°\mathrm{F}) - 32)$
Celsius to Kelvin	$T(K) = T(°\mathrm{C}) + 273.15$
Kelvin to Celsius	$T(°\mathrm{C}) = T(K) - 273.15$
Fahrenheit to Kelvin	$T(K) = \frac{5}{9}(T(°\mathrm{F}) - 32) + 273.15$
Kelvin to Fahrenheit	$T(°\mathrm{F}) = \frac{9}{5}(T(K) - 273.15) + 32$

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. This extrapolation implies that there is a lowest temperature. This temperature is called absolute zero. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is – 273.15 ºC or 0 K.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the *zeroth law of thermodynamics*. This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the zeroth law because it comes logically before the first and second laws discussed in Thermodynamics.

What is the underlying cause of thermal expansion? As is discussed in *Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature*, an increase in temperature implies an increase in the kinetic energy of the individual atoms. In a solid, unlike in a gas, the atoms or molecules are closely packed together, but their kinetic energy (in the form of small, rapid vibrations) pushes neighboring atoms or molecules apart from each other. This neighbor-to-neighbor pushing results in a slightly greater distance, on average, between neighbors, and adds up to a larger size for the whole body. For most substances under ordinary conditions, there is no preferred direction, and an increase in temperature will increase the solid's size by a certain fraction in each dimension. The change in length ΔL is proportional to length L. The dependence of thermal expansion on temperature, substance, and length is summarized in the equation

$$\Delta L = \alpha L \Delta T, \tag{1.2.1}$$

where ΔL is the change in length L, ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an *ideal gas*, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the *ideal gas law*. The ideal gas law states that

$$PV = NkT, \tag{1.2.2}$$

where P (pressure is the force divided by the area on which the force is exerted) is the absolute pressure of a gas [Pa], V is the volume it occupies [m^3], N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the *Boltzmann constant* in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value $k = 1.38 \cdot 10^{-23}$ [J/K]. It is sometimes convenient to work with a unit other than molecules when measuring the amount of substance. A mole abbreviated [mol] is defined to be the amount of a substance that contains as many atoms or molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. The actual number of atoms or molecules in one mole is called *Avogadro's number* (N_A), in recognition of Italian scientist Amedeo Avogadro (1776–1856). He developed the concept of the mole, based on the hypothesis that equal volumes of gas, at the same pressure and temperature, contain equal numbers of molecules. That is, the number is independent of the type of gas. This hypothesis has been confirmed, and the value of Avogadro's number is $N_A = 6.02 \cdot 10^{23}$ [mol^{-1}].

Because a huge number of molecules will collide with the wall in a short time, we observe an average force per unit area. These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions and thus the pressure increase. The following relationship is found:

$$PV = \frac{1}{3}Nm\overline{v^2}, \tag{1.2.3}$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

We can get the average kinetic energy of a molecule, $\frac{1}{2}mv^2$, from the left-hand side of the equation by canceling N and multiplying by 3/2. This calculation produces the result that the average kinetic energy of a molecule is directly related to absolute temperature:

$$\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT. \tag{1.2.4}$$

The average translational kinetic energy of a molecule $\overline{KE}$, is called *thermal energy*. And solve for the average speed of molecules in a gas in terms of temperature:

$$\sqrt{\overline{v^2}} = v_{rms} = \sqrt{\frac{3kT}{m}}, \tag{1.2.5}$$

where v_{rms} stands for root-mean-square (rms) speed.

The kinetic theory of gases was developed by Daniel Bernoulli (1700–1782), who is best known in physics for his work on fluid flow (hydrodynamics). Bernoulli's work predates the atomistic view of matter established by Dalton. The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This distribution is called the Maxwell-Boltzmann distribution, after its originators, who calculated it based on kinetic theory, and has since been confirmed experimentally (Fig. 4). The distribution has a long tail, because a few molecules may go several times the rms speed. The most probable speed v_p is less than the rms speed v_{rms}.

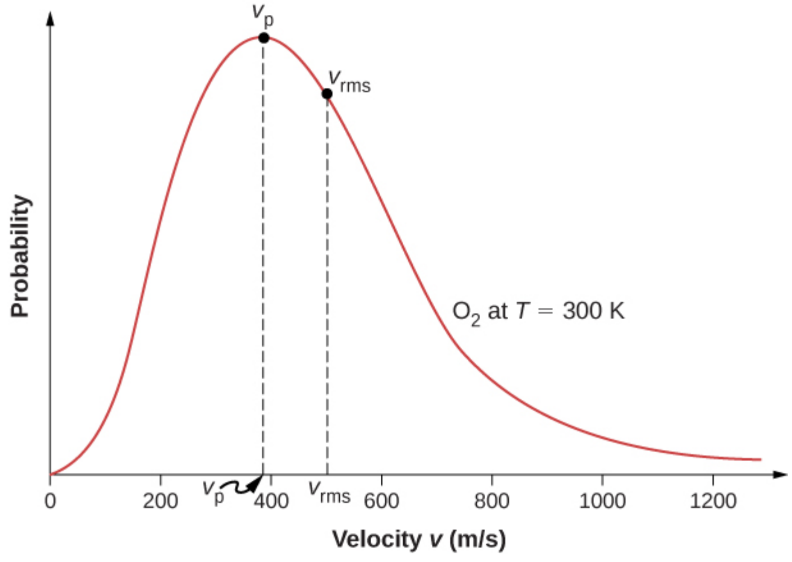

Fig. 4. The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas

Real gases are like ideal gases at high temperatures. At lower temperatures, however, the interactions between the molecules and their volumes cannot be ignored. The molecules are very close (condensation occurs) and there is a dramatic decrease in volume. The plots of pressure versus temperatures provide considerable insight into thermal properties of substances. There are well-defined regions on these graphs that correspond to various phases of matter, so PT-graphs are called *phase diagrams*. Fig. 5 shows the phase diagram for water.

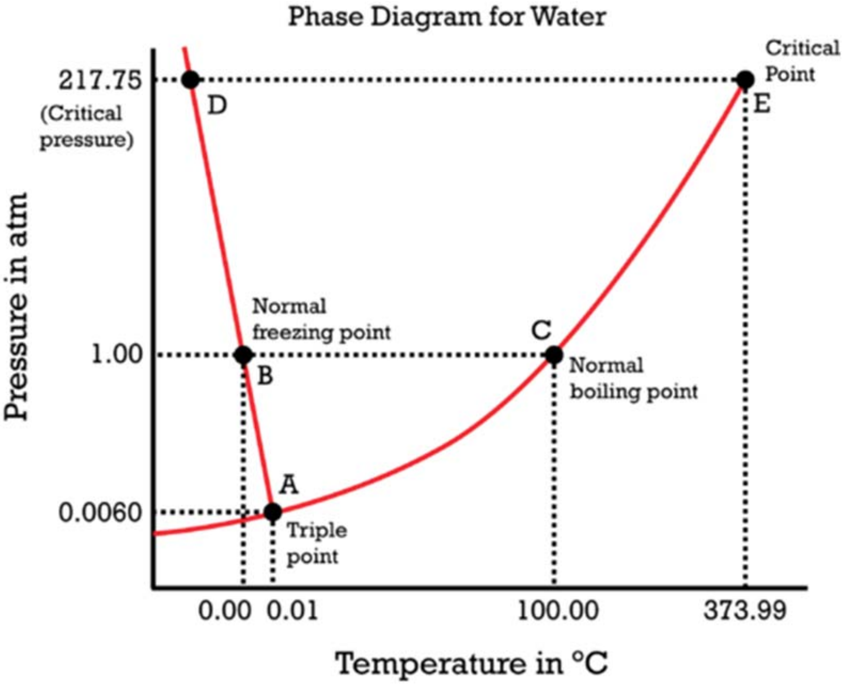

Fig. 5. The phase diagram (PT-graph) for water

Using the graph, if you know the pressure and temperature you can determine the phase of water. The solid lines–boundaries between phases–indicate temperatures and pressures at which the phases coexist (that is, they exist together in ratios, depending on pressure and temperature). For example, the boiling point of water is 100 ºC at 1.00 atm. As the pressure increases, the boiling temperature rises steadily to 374 ºC at a pressure of 218 atm. The curve ends at a point called the critical point, because at higher temperatures the liquid phase does not exist at any pressure. The critical point occurs at the critical temperature. The critical temperature for oxygen (O_2) is –118 ºC, so oxygen cannot be liquefied above this temperature. All three curves on the phase diagram meet at a single point, the triple point, where all three phases exist in equilibrium. For water, the triple point occurs at 273.16 K (0.01 ºC), and is a more accurate calibration temperature than the melting point of water at 1.00 atm, or 273.15 K (0.0 ºC).

Liquid and gas phases are in equilibrium at the boiling temperature. If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their relative amount. Molecules in the liquid escape as a gas at the same rate at which gas molecules stick to the liquid, or form droplets and become part of the liquid phase. The combination of temperature and pressure has to be “just right”; if the temperature and pressure are increased, equilibrium is maintained by the same increase of boiling and condensation rates.

Vapor pressure is defined as the pressure at which a gas coexists with its solid or liquid phase. Vapor pressure is created by faster molecules that break away from the liquid or solid and enter the gas phase. The vapor pressure of a substance depends on both the substance and its temperature – an increase in temperature increases the vapor pressure. *Partial pressure* is defined as the pressure a gas would create if it occupied the total volume available. In a mixture of gases, the total pressure is the sum of partial pressures of the component gases, assuming ideal gas behavior and no chemical reactions between the components. This law is known as *Dalton’s law of partial pressures*, after the English scientist John Dalton (1766–1844), who proposed it. Dalton’s law is based on kinetic theory, where each gas creates its pressure by molecular collisions, independent of other gases present. It is consistent with the fact that pressures add according to *Pascal’s Principle*. Thus water evaporates and ice sublimates when their vapor pressures exceed the partial pressure of water vapor in the surrounding mixture of gases. If their vapor pressures are less than the partial pressure of water vapor in the surrounding gas, liquid droplets or ice crystals (frost) form.

When we say humidity, we really mean *relative humidity*. Relative humidity tells us how much water vapor is in the air compared with the maximum possible. At its maximum, denoted as *saturation*, the relative humidity is 100 %, and evaporation is inhibited. The amount of water vapor the air can hold depends on its temperature. At the *dew point* temperature, relative humidity is 100 %, and fog may result from the condensation of water droplets if they are small enough to stay in suspension. We define percent relative humidity as the ratio of vapor density to saturation vapor density, or

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100. \tag{1.2.6}$$

We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i. e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference, and ceases once the temperatures are equal. These observations lead to the following definition of heat: *Heat* – is the spontaneous transfer of energy due to a temperature difference. Owing to the fact that heat is a form of energy, it has the SI unit of joule [J]. The calorie [cal] is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00 ºC – specifically, between 14.5 ºC and 15.5 ºC, since there is a slight temperature dependence. Perhaps the most common unit of heat is the kilocalorie [kcal], which is the energy needed to change the temperature of 1.00 kg of water by 1.00 ºC. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called "big calorie") are actually kilocalories (1 kilocalorie = 1000 calories), a fact not easily determined from package labeling. It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the *mechanical equivalent of heat* – the work needed to produce the same effects as heat transfer. In terms of the units used for these two terms, the best modern value for this equivalence is 1.000 [kcal] = 4186 [J]. We consider this equation as the conversion between two different units of energy.

The quantitative relationship between heat transfer and temperature change contains all three factors:

$$Q = mcdT, \tag{1.2.7}$$

where Q is the symbol for heat transfer, m is the mass of the substance, and dT is the change in temperature. The symbol c stands for specific heat and depends on the material and phase. The specific heat is the amount of heat necessary to change the temperature of 1.00 kg of mass by 1.00 ºC. The specific heat c is a property of the substance; its SI unit is [J/(kg · K] or [J/kg · ºC]. Recall that the temperature change (dT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then the unit of specific heat is [kcal/kg · ºC].

The heat Q required to change the phase of a sample of mass m is given by

$$Q = mL_f \quad \text{or} \quad Q = mL_v, \tag{1.2.8}$$

where the latent heat of fusion, Lf, and latent heat of vaporization, Lv, are material constants that are determined experimentally.

Every process involving heat transfer takes place by only three methods:

I. *Conduction* is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale–we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.

II. *Convection* is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.

III. Heat transfer by *radiation* occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.

Basic physical laws govern how heat transfer for doing work takes place and place insurmountable limits onto its efficiency. If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The *first law of thermodynamics states* that the change in internal energy of a system equals the net heat transfer into the system minus the net work done by the system. In equation form, the first law of thermodynamics is:

$$dU = \delta Q - \delta W. \tag{1.2.9}$$

Here dU is the change in internal energy U of the system. δQ is the net heat transferred into the system – that is, δQ is the sum of all heat transfer into and out of the system. δW is the net work done by the system – that is, δW is the sum of all work done on or by the system. We use the following sign conventions: if δQ is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive δQ adds energy to the system and positive δW takes energy from the system. Thus (1.2.9) note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this – heat transfer into them takes place so that they can do work. The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

A process by which a gas does work on a piston at constant pressure is called an *isobaric process*. Since the pressure is constant, the force exerted is constant and the work done is given as:

$$W = P\Delta V. \tag{1.2.10}$$

Volume is constant, $\Delta V = 0$, and no work is done in an *isochoric process*

$$W = 0. \tag{1.2.11}$$

Isothermal process – that is, one in which temperature is kept constant

$$Q = W. \tag{1.2.12}$$

An *adiabatic process*, defined to be one in which there is no heat transfer – that is

$$Q = 0. \tag{1.2.13}$$

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur – none of those processes violate conservation of energy. The law that forbids these processes is called the *second law of thermodynamics*. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes. The already familiar direction of heat transfer from hot to cold is the basis of our first version of the second law of thermodynamics: Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction. The second law of thermodynamics also states, with regard to using heat transfer to do work (the second expression of the second law): It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state. *Cyclical process* brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Thus the net work done by the system equals the net heat transfer into the system, or

$$W = Q_h - Q_c, \tag{1.2.14}$$

Q_h is the heat transfer out of the hot reservoir, W is the work output, and Q_c is the heat transfer into the cold reservoir.

We know from the second law of thermodynamics that a heat engine cannot be 100 % efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the *Carnot cycle*, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a *Carnot engine*. What is crucial to the Carnot cycle – and, in fact, defines it – is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior. Stated in terms of reversible processes, the second law of thermodynamics has a third form: A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

Carnot also determined the efficiency of a perfect heat engine–that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

$$E_{ff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}. \tag{1.2.15}$$

The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a *heat pump's coefficient of performance* (COP_{hp}) to be

$$COP_{hp} = \frac{Q_h}{W}. \tag{1.2.16}$$

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called *entropy*. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work. This ratio of Q/T is defined to be the *change in entropy* dS for a reversible process

$$dS = \frac{dQ}{T}, \tag{1.2.17}$$

where dQ is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin [J/K]. If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find *dS*. It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is *dQ/T*, there is a larger change at lower temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase. This result is very general: There is an increase in entropy for any system undergoing an irreversible process. With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of the *second law of thermodynamics stated in terms of entropy*: The total entropy of a system either increases or remains constant in any process; it never decreases.

1.3. Electricity and Magnetism

Atomic and molecular interactions, such as the forces of friction, cohesion, and adhesion, are now known to be manifestations of the *electromagnetic force*. *Static electricity* is just one aspect of the electromagnetic force, which also includes moving electricity and magnetism. All the macroscopic forces that we experience directly, such as the sensations of touch and the tension in a rope, are due to the electromagnetic force, one of the four fundamental forces in nature. The gravitational force, another fundamental force, is actually sensed through the electromagnetic interaction of molecules, such as between those in our feet and those on the top of a bathroom scale. The other two fundamental forces, the strong nuclear force and the weak nuclear force, cannot be sensed on the human scale. Some of the most basic characteristics of static electricity include:

- The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge.
- There are only two types of charge, one called positive and the other called negative.
- Like charges repel, whereas unlike charges attract.
- The force between charges decreases with distance.

Fig. 6 shows a simple model of an atom with negative electrons orbiting its positive nucleus. The nucleus is positive due to the presence of positively charged protons. Nearly all charge in nature is due to electrons and protons, which are two of the three building blocks of most matter. The third is the neutron, which is neutral, carrying no charge. Other charge-carrying particles are observed in cosmic rays and nuclear decay, and are created in particle accelerators. All but the electron and proton survive only a short time and are quite rare by comparison.

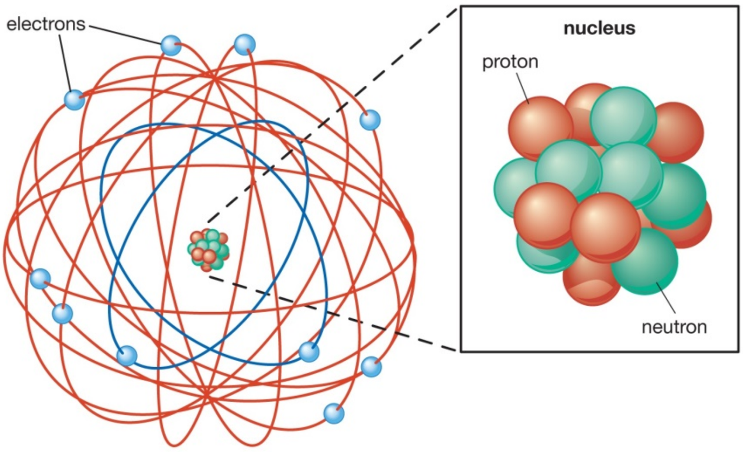

Fig. 6. The effects of static electricity are explained by a physical quantity not previously introduced, called electric charge

The charges of electrons and protons are identical in magnitude but opposite in sign. Furthermore, all charged objects in nature are integral multiples of this basic quantity of charge, meaning that all charges are made of combinations of a basic unit of charge. Usually, charges are formed by combinations of electrons and protons. The magnitude of this basic charge is

$$|q_e| = 1.602176487 \cdot 10^{-19}[C]. \tag{1.3.1}$$

The symbol *q* is commonly used for charge and the subscript *e* indicates the charge of a single electron (or proton). The SI unit of charge is the coulomb [C]. In fact, in all situations the total amount of charge is always constant. This universally obeyed law of nature is called *the law of conservation of charge*: total charge is constant in any process. Only a limited number of physical quantities are universally conserved. Charge is one – energy, momentum, and angular momentum are others. Because they are conserved, these physical quantities are used to explain more phenomena and form more connections than other, less basic quantities. We find that conserved quantities give us great insight into the rules followed by nature and hints to the organization of nature. Discoveries of conservation laws have led to further discoveries, such as the weak nuclear force and the quark substructure of protons and other particles.

Some substances, such as metals and salty water, allow charges to move through them with relative ease. Some of the electrons in metals and similar conductors are not bound to individual atoms or sites in the material. These free electrons can move through the material much as air moves through loose sand. Any substance that has free electrons and allows charge to move relatively freely through it is called a conductor. The moving electrons may collide with fixed atoms and molecules, losing some energy, but they can move in a conductor. Superconductors allow the movement of charge without any loss of energy. Salty water and other similar conducting materials contain free ions that can move through them. An ion is an atom or molecule having a positive or negative (nonzero) total charge. In other words, the total number of electrons is not equal to the total number of protons.

Through the work of scientists in the late 18th century, the main features of the electrostatic force – the existence of two types of charge, the observation that like charges repel, unlike charges attract, and the decrease of force with distance – were eventually refined, and expressed as a mathematical formula. The mathematical formula for the electrostatic force is called *Coulomb's law* after the French physicist Charles Coulomb (1736–1806), who performed experiments and first proposed a formula to calculate it.

$$F = k\frac{|q_1 q_2|}{r^2}. \tag{1.3.2}$$

Coulomb's law calculates the magnitude of the force F between two point charges, q_1 and q_2, separated by a distance r. In SI units, the constant k is equal to

$$k \approx 8.99 \cdot 10^9 \, [N \cdot m^2 \cdot C^{-2}]. \tag{1.3.3}$$

The electrostatic force is a vector quantity and is expressed in units of newtons. The force is understood to be along the line joining the two charges. A field is a way of conceptualizing and mapping the force that surrounds any object and acts on another object at a distance without apparent physical connection. For example, the gravitational field surrounding the earth (and all other masses) represents the gravitational force that would be experienced if another mass were placed at a given point within the field. In the same way, the Coulomb force field surrounding any charge extends throughout space. Using Coulomb's law, its magnitude is given by the equation $F = k\,|qQ|/r^2$, for a point charge (a particle having a charge Q) acting on a test charge q at a distance r. Both the magnitude and direction of the Coulomb force field depend on Q and the test charge q.

To simplify things, we would prefer to have a field that depends only on Q and not on the test charge q. The electric field is defined in such a manner that it represents only the charge creating it and is unique at every point in space. Specifically, the electric field E is defined to be the ratio of the Coulomb force to the test charge:

$$E = \frac{F}{q}, \tag{1.3.4}$$

where F is the electrostatic force (or Coulomb force) exerted on a positive test charge q. It is understood that E is in the same direction as F. It is also assumed that q is so small that it does not alter the charge distribution creating the electric field. The units of electric field are newtons per coulomb [*N/C*]. If the electric field is known, then the electrostatic force on any charge q is simply obtained by multiplying charge times electric field, or $F = qE$. Consider the electric field due to a point charge Q. According to Coulomb's law, the force it exerts on a test charge q is $F = k\,|qQ|/r^2$. Thus the magnitude of the electric field E, for a point charge is

$$E = \left|\frac{F}{q}\right| = k\left|\frac{qQ}{qr^2}\right| = k\left|\frac{Q}{r^2}\right|. \tag{1.3.5}$$

The electric field is thus seen to depend only on the charge Q and the distance r; it is completely independent of the test charge q. Drawings using lines to represent electric fields around charged objects are very useful in visualizing field strength and direction. Since the electric field has both magnitude and direction, it is a vector. Like all vectors, the electric field can be represented by an arrow that has length proportional to its magnitude and that point in the correct direction. Fig. 7 shows two pictorial representations of the same electric field created by a positive point charge *Q*. Fig. 7, *b* shows the standard representation using continuous lines. Fig. 7, *a* shows numerous individual arrows with each arrow representing the force on a test charge *q*. Field lines are essentially a map of infinitesimal force vectors.

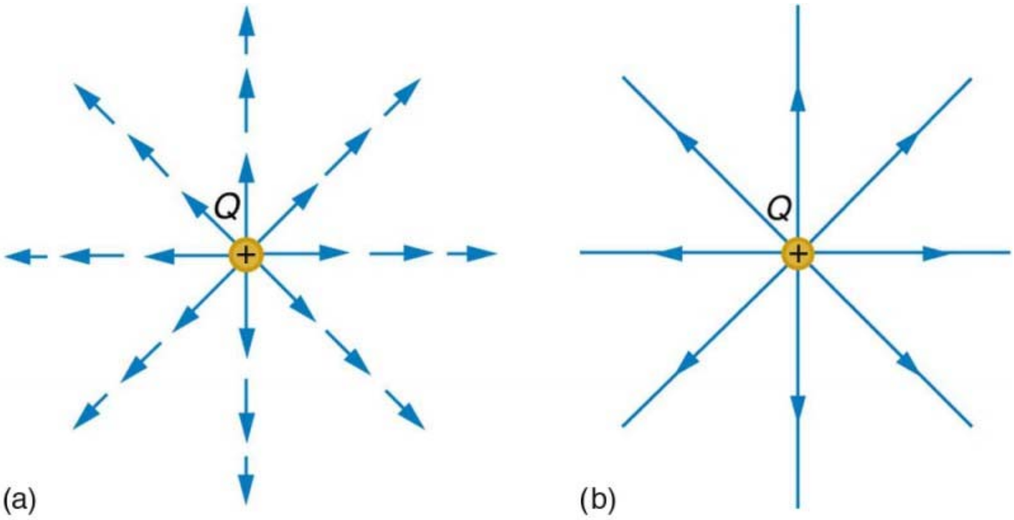

Fig. 7. Two equivalent representations of the electric field due to a positive charge *Q*: *a* – arrows representing the electric field's magnitude and direction; *b* – in the standard representation, the arrows are replaced by continuous field lines having the same direction at any point as the electric field

We use electric field lines to visualize and analyze electric fields (the lines are a pictorial tool, not a physical entity in themselves). The properties of electric field lines for any charge distribution can be summarized as follows:

1. Field lines must begin on positive charges and terminate on negative charges, or at infinity in the hypothetical case of isolated charges.
2. The number of field lines leaving a positive charge or entering a negative charge is proportional to the magnitude of the charge.
3. The strength of the field is proportional to the closeness of the field lines – more precisely, it is proportional to the number of lines per unit area perpendicular to the lines.
4. The direction of the electric field is tangent to the field line at any point in space.
5. Field lines can never cross.

The last property means that the field is unique at any point. The field line represents the direction of the field; so if they crossed, the field would have two directions at that location (an impossibility if the field is unique).

Conductors contain *free charges* that move easily. When excess charge is placed on a conductor or the conductor is put into a static electric field, charges in the conductor quickly respond to reach a steady state called *electrostatic equilibrium*. A conductor placed in an electric field will be polarized. Properties of a Conductor in Electrostatic Equilibrium:

1. The electric field is zero inside a conductor.
2. Just outside a conductor, the electric field lines are perpendicular to its surface, ending or beginning on charges on the surface.
3. Any excess charge resides entirely on the surface or surfaces of a conductor.

The study of electrostatics has proven useful in many areas. Van de Graaff generators (or Van de Graaffs) are not only spectacular devices used to demonstrate high voltage due to static electricity – they are also used for serious research. The first was built by Robert Van de Graaff in 1931 (based on original suggestions by Lord Kelvin) for use in nuclear physics research. Van de Graaffs utilize both smooth and pointed surfaces, and conductors and insulators to generate large static charges and, hence, large voltages. Two of the most familiar aspects of electricity are its *energy* and *voltage*. We know, for example, that great amounts of electrical energy can be stored in batteries, are transmitted cross-country through power lines, and may jump from clouds to explode the sap of trees. In a similar manner, at molecular levels, ions cross cell membranes and transfer information. We also know about voltages associated with electricity. Batteries are typically a few volts, the outlets in your home produce 120 volts, and power lines can be as high as hundreds of thousands of volts. But energy and voltage are not the same thing. We use the letters PE to denote electric potential energy, which has units of joules [J]. The change in potential energy, ΔPE, is crucial, since the work done by a conservative force is the negative of the change in potential energy; that is, $W = -\Delta PE$. For example, work W done to accelerate a positive charge from rest is positive and results from a loss in PE, or a negative ΔPE. There must be a minus sign in front of ΔPE to make W positive. PE can be found at any point by taking one point as a reference and calculating the work needed to move a charge to the other point.

Gravitational potential energy and electric potential energy are quite analogous. Potential energy accounts for work done by a conservative force and gives added insight regarding energy and energy transformation without the necessity of dealing with the force directly. It is much more common, for example, to use the concept of voltage (related to electric potential energy) than to deal with the Coulomb force directly. Calculating the work directly is generally difficult, since $W = Fd\cos\theta$ and the direction and magnitude of F can be complex for multiple charges, for odd-shaped objects, and along arbitrary paths. But we do know that, since $F = qE$, the work, and hence ΔPE, is proportional to the test charge q. To have a physical quantity that is independent of test charge, we define electric potential V (or simply potential, since electric is understood) to be the potential energy per unit charge:

$$V = \frac{PE}{q}. \tag{1.3.6}$$

The potential difference between points 1 and 2, $V_2 - V_1$, is thus defined to be the change in potential energy of a charge q moved from 1 to 2, divided by the charge. Units of potential difference are joules per coulomb, given the name volt [V] after Alessandro Volta. The familiar term *voltage* is the common name for potential difference. Keep in mind that whenever a voltage is quoted, it is understood to be the potential difference between two points. For example, every battery has two terminals, and its voltage is the potential difference between them. More fundamentally, the point you choose to be zero volts is arbitrary. This is analogous to the fact that gravitational potential energy has an arbitrary zero, such as sea level or perhaps a lecture hall floor.

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

$$1eV = 1.602176634 \cdot 10^{-19}[\mathrm{J}]. \tag{1.3.7}$$

Point charges, such as electrons, are among the fundamental building blocks of matter. Furthermore, spherical charge distributions (like on a metal sphere) create external electric fields exactly like a point charge. The electric potential due to a point charge is, thus, a case we need to consider. Using calculus to find the work needed to move a test charge q from a large distance away to a distance of r from a point charge Q, and noting the connection between work and potential, it can be shown that the electric potential V of a point charge is

$$V = \frac{kQ}{r} \text{ (Point Charge).} \tag{1.3.8}$$

We can represent electric potentials (voltages) pictorially, just as we drew pictures to illustrate electric fields. Of course, the two are related.

A *capacitor* is a device used to store electric charge. Capacitors have applications ranging from filtering static out of radio reception to energy storage in heart defibrillators. Typically, commercial capacitors have two conducting parts close to one another, but not touching. When battery terminals are connected to an initially uncharged capacitor, equal amounts of positive and negative charge, $+Q$ and $–Q$, are separated into its two plates. The capacitor remains neutral overall, but we refer to it as storing a charge Q in this circumstance. The amount of charge Q a capacitor can store depends on two major factors – the voltage applied and the capacitor's physical characteristics, such as its size. Different *capacitors* will store different amounts of charge for the same applied voltage, depending on their physical characteristics. We define their capacitance C to be such that the charge Q stored in a capacitor is proportional to C. The charge stored in a capacitor is given by

$$Q = CV. \tag{1.3.9}$$

The unit of capacitance is the farad [F], named for Michael Faraday (1791–1867), an English scientist who contributed to the fields of electromagnetism and electrochemistry. Since capacitance is charge per unit voltage, we see that a farad is a coulomb per volt, or [1 F = 1 C/1V]. It can be shown that for a parallel plate capacitor there are only two factors (A and d) that affect its capacitance C. The capacitance of a parallel plate capacitor in equation form is given by

$$C = \varepsilon_0 \frac{A}{d}. \tag{1.3.10}$$

The previous example highlights the difficulty of storing a large amount of charge in capacitors. If d is made smaller to produce a larger capacitance, then the maximum voltage must be reduced proportionally to avoid breakdown. An important solution to this difficulty is to put an insulating material, called a dielectric, between the plates of a capacitor and allow d to be as small as possible. Not only does the smaller d make the capacitance greater, but many insulators can withstand greater electric fields than air before breaking down. There is another benefit to using a dielectric in a capacitor. Depending on the material used, the capacitance is greater than that given by the equation (1.3.10) by a factor ε, called the *dielectric constant.* A parallel plate capacitor with a dielectric between its plates has a capacitance given by

$$C = \varepsilon\varepsilon_0 \frac{A}{d}. \tag{1.3.11}$$

Values of the dielectric constant ε for various materials are given in Table 6. Note that ε for vacuum is exactly 1, and so the above equation is valid in that case, too.

Table 6

Dielectric Constants and Dielectric Strengths for Various Materials at 20 ºC

Vacuum	1.00000	–
Air	1.00059	$3 \cdot 10^6$
Paper	3.7	$16 \cdot 10^6$
Pyrex Glass	5.6	$14 \cdot 10^6$
Teflon	2.1	$60 \cdot 10^6$
Water	80	–

Total capacitance in series, C_s

$$\frac{1}{C_s} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots \tag{1.3.12}$$

Total capacitance in parallel, C_p

$$C_p = C_1 + C_2 + C_3 + \cdots \tag{1.3.13}$$

The energy stored in a capacitor can be expressed in three ways:

$$E_{cap} = \frac{QV}{2} = \frac{CV^2}{2} = \frac{Q^2}{2C}, \tag{1.3.14}$$

where Q is the charge, V is the voltage, and C is the capacitance of the capacitor. The energy is in joules for a charge in coulombs, voltage in volts, and capacitance in farads.

Electric current is defined to be the rate at which charge flows. A large current, such as that used to start a truck engine, moves a large amount of charge in a small time, whereas a small current, such as that used to operate a hand-held calculator, moves a small amount of charge over a long period of time. In equation form, *electric current I* is defined to be

$$I = \frac{dQ}{dt}, \tag{1.3.15}$$

where dQ is the amount of charge passing through a given area in time dt. The SI unit for current is the ampere [A], named for the French physicist André-Marie Ampère (1775–1836). Since (1.3.15), we see that an ampere is one coulomb per second [1 A = 1 C/1s].

The current that flows through most substances is directly proportional to the voltage V applied to it. The German physicist Georg Simon Ohm (1787–1854) was the first to demonstrate experimentally that the current in a metal wire is directly proportional to the voltage applied. This important relationship is known as *Ohm's law*. It can be viewed as a cause-and-effect relationship, with voltage the cause and current the effect. This is an empirical law like that for friction – an experimentally observed phenomenon. Such a linear relationship doesn't always occur. The electric property that impedes current (crudely similar to friction and air resistance) is called *resistance R*. Collisions of moving charges with atoms and molecules in a substance transfer energy to the substance and limit current. Resistance is defined as inversely proportional to current, or

$$I = \frac{V}{R}. \tag{1.3.16}$$

The many substances for which Ohm's law holds are called *ohmic*. These include good conductors like copper and aluminum, and some poor conductors under certain circumstances. Ohmic materials have a resistance R that is independent of voltage V and current I. An object that has simple resistance is called a resistor, even if its resistance is small. The unit for resistance is an ohm and is given the symbol Ω (upper case Greek omega).

The resistance of an object depends on its shape and the material of which it is composed. The cylindrical resistor in Fig. 8 is easy to analyze, and, by so doing, we can gain insight into the resistance of more complicated shapes. As you might expect, the cylinder's electric resistance R is directly proportional to its length L, similar to the resistance of a pipe to fluid flow. The longer the cylinder, the more collisions charges will make with its atoms. The greater the diameter of the cylinder, the more current it can carry. In fact, R is inversely proportional to the cylinder's cross-sectional area A:

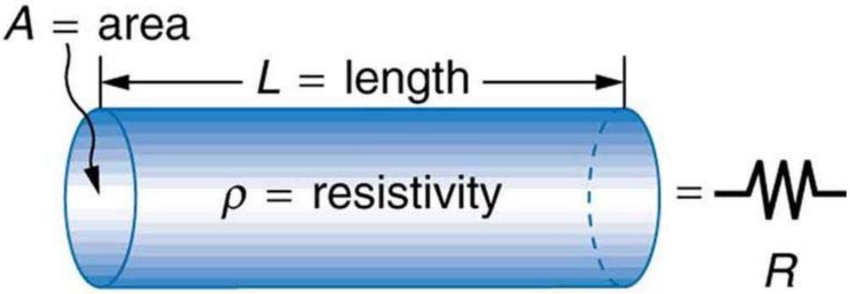

Fig. 8. A uniform cylinder of length L and cross-sectional area A. Its resistance to the flow of current is similar to the resistance posed by a pipe to fluid flow

For a given shape, the resistance depends on the material of which the object is composed. Different materials offer different resistance to the flow of charge. We define the *resistivity* ρ of a substance so that the resistance R of an object is directly proportional to ρ. Resistivity ρ is an intrinsic property of a material, independent of its shape or size. The resistance R of a uniform cylinder of length L, of cross-sectional area A, and made of a material with resistivity ρ, is

$$R = \frac{\rho L}{A}. \tag{1.3.17}$$

Table 7 gives representative values of ρ.

Table 7

Resistivities ρ of various materials at 20 ºC

Material	ρ [Ω · m]
Silver	$1.59 \cdot 10^{-8}$
Cooper	$1.79 \cdot 10^{-8}$
Gold	$2.44 \cdot 10^{-8}$
Glass	$10^{9} – 10^{14}$
Wood	$10^{8} – 10^{11}$

The resistivity of all materials depends on temperature. Some even become superconductors (zero resistivity) at very low temperatures. Conversely, the resistivity of conductors increases with increasing temperature. Since the atoms vibrate more rapidly and over larger distances at higher temperatures, the electrons moving through a metal make more collisions, effectively making the resistivity higher. Over relatively small temperature changes (about 100 ºC or less), resistivity ρ varies with temperature change ΔT as expressed in the following equation

$$\rho = \rho_0(1 + \alpha\Delta T), \tag{1.3.18}$$

where ρ_0 is the original resistivity and α is the *temperature coefficient of resistivity*. For larger temperature changes, α may vary or a nonlinear equation may be needed to find ρ. Note that α is positive for metals, meaning their resistivity increases with temperature. Some alloys have been developed specifically to have a small temperature dependence.

Electric energy depends on both the voltage involved and the charge moved. This is expressed most simply as $PE = qV$, where q is the charge moved and V is the voltage (or more precisely, the potential difference the charge moves through). Power is the rate at which energy is moved, and so electric power is

$$P = \frac{PE}{t} = \frac{qV}{t}. \tag{1.3.19}$$

Electric power P is simply the product of current times voltage. Power has familiar units of watts. Since the SI unit for potential energy *PE* is the joule, power has units of joules per second, or watts. Thus, [1A · V = 1 W]. Three expressions for electric power are listed together here for convenience

$$P = IV = \frac{V^2}{R} = I^2R. \tag{1.3.20}$$

Most of the examples dealt with so far, and particularly those utilizing batteries, have constant voltage sources. Once the current is established, it is thus also a constant. *Direct current* (DC) is the flow of electric charge in only one direction. It is the steady state of a constant-voltage circuit. Most well-known applications, however, use a time-varying voltage source. *Alternating current* (AC) is the flow of electric charge that periodically reverses direction. If the source varies periodically, particularly sinusoidally, the circuit is known as an alternating current circuit. Examples include the commercial and residential power that serves so many of our needs. Fig. 9 shows graphs of voltage and current versus time for typical DC and AC power. The AC voltages and frequencies commonly used in homes and businesses vary around the world.

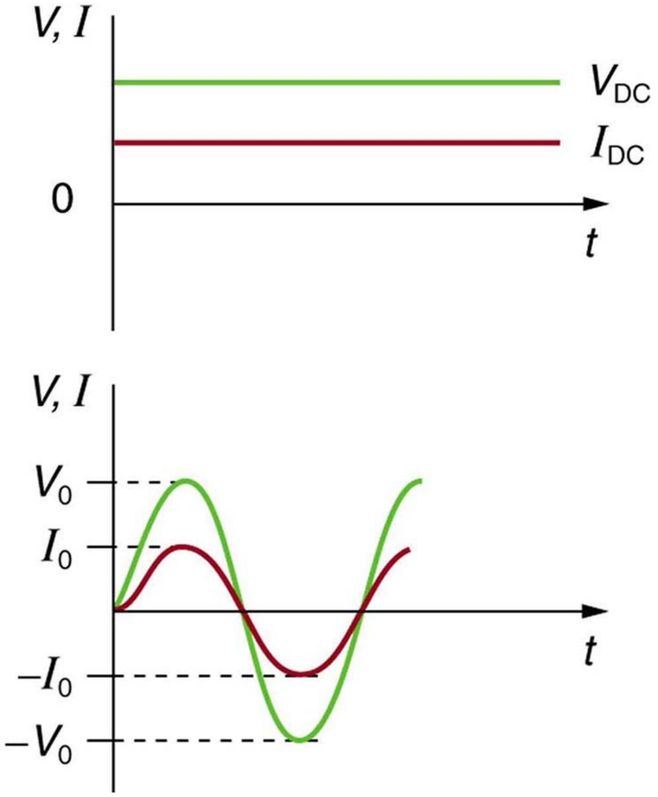

Fig. 9. a – DC voltage and current are constant in time, once the current is established; *b* – a graph of voltage and current versus time for 60-Hz AC power. The voltage and current are sinusoidal and are in phase for a simple resistance circuit. The frequencies and peak voltages of AC sources differ greatly

Most circuits have more than one component, called a *resistor* that limits the flow of charge in the circuit. A measure of this limit on charge flow is called *resistance*. The simplest combinations of resistors are the series and parallel connections illustrated in Fig. 10. The total resistance of a combination of resistors depends on both their individual values and how they are connected.

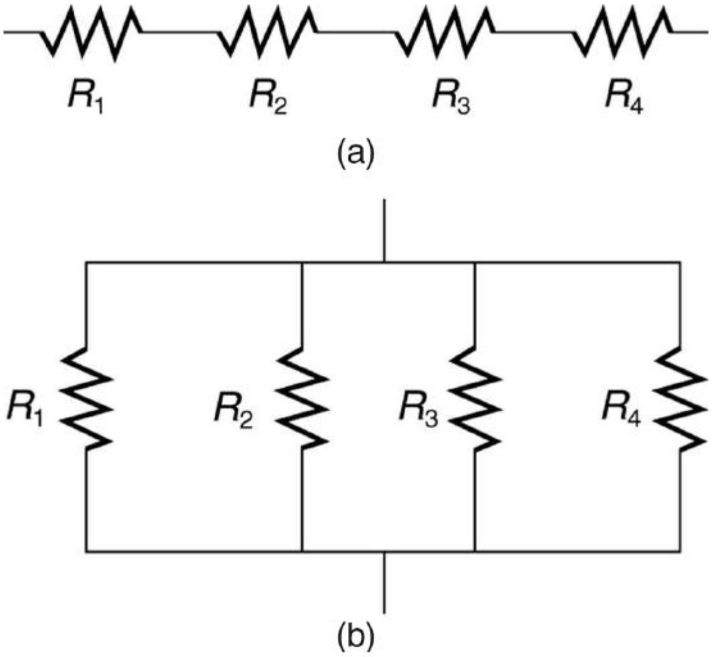

Fig. 10. a – a series connection of resistors; *b* – a parallel connection of resistors

This implies that the total or equivalent series resistance R_s of four resistors is $R_S = R_1 + R_2 + R_3 + R_4$. This logic is valid in general for any number of resistors in series; thus, the total resistance R_s of a series connection is

$$R_S = R_1 + R_2 + R_3 + \cdots, \tag{1.3.21}$$

as proposed. Since all of the current must pass through each resistor, it experiences the resistance of each, and resistances in series simply add up.

The total resistance R_p of a parallel connection is related to the individual resistances by

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \tag{1.3.22}$$

This relationship results in a total resistance R_p that is less than the smallest of the individual resistances.

Many complex circuits, such as the one in Fig. 11, cannot be analyzed with the series-parallel techniques developed. There are, however, two circuit analysis rules that can be used to analyze any circuit, simple or complex. These rules are special cases of the laws of conservation of charge and conservation of energy. The rules are known as *Kirchhoff's rules*, after their inventor Gustav Kirchhoff (1824–1887).

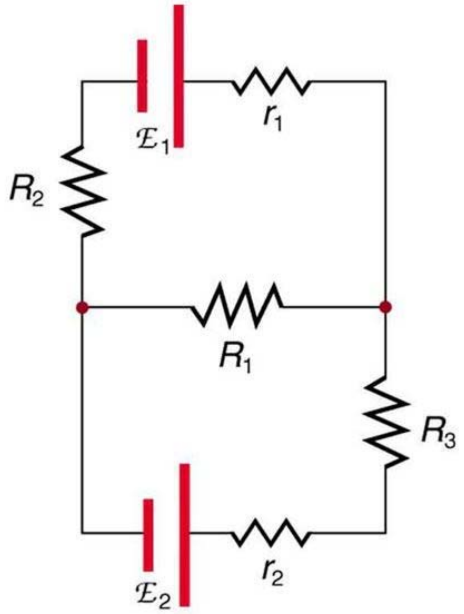

Fig. 11. Complex circuits. Note: the script E in the figure represents electromotive force, emf

- Kirchhoff's first rule – the junction rule. The sum of all currents entering a junction must equal the sum of all currents leaving the junction.
- Kirchhoff's second rule – the loop rule. The algebraic sum of changes in potential around any closed circuit path (loop) must be zero.

Another important part of the electromagnetism is magnetism. All magnets attract iron, such as that in a refrigerator door. However, magnets may attract or repel other magnets. Experimentation shows that all magnets have two poles. If freely suspended, one pole will point toward the north. The two poles are thus named the *north magnetic pole* and the *south magnetic pole* (or more properly, north-seeking and south-seeking poles, for the attractions in those directions). The Earth acts like a very large bar magnet with its south-seeking pole near the geographic North Pole. That is why the north pole of your compass is attracted toward the geographic north pole of the Earth – because the magnetic pole that is near the geographic North Pole is actually a south magnetic pole! Confusion arises because the geographic term "North Pole" has come to be used (incorrectly) for the magnetic pole that is near the North Pole. Thus, "North magnetic pole" is actually a misnomer – it should be called the South magnetic pole.

Only certain materials, such as iron, cobalt, nickel, and gadolinium, exhibit strong magnetic effects. Such materials are called *ferromagnetic*, after the Latin word for iron, ferrum. A group of materials made from the alloys of the rare earth elements are also used as strong and permanent magnets; a popular one is neodymium. Other materials exhibit weak magnetic effects, which are detectable only with sensitive instruments. Not only do ferromagnetic materials respond strongly to magnets (the way iron is attracted to magnets), they can also be *magnetized* themselves – that is, they can be induced to be magnetic or made into permanent magnets.

Early in the 19th century, it was discovered that electrical currents cause magnetic effects. The first significant observation was by the Danish scientist Hans Christian Oersted (1777–1851), who found that a compass needle was deflected by a current-carrying wire. This was the first significant evidence that the movement of charges had any connection with magnets. *Electromagnetism* is the use of electric current to make magnets. These temporarily induced magnets are called *electromagnets*.

Since magnetic forces act at a distance, we define a *magnetic field* to represent magnetic forces. The pictorial representation of *magnetic field lines* is very useful in visualizing the strength and direction of the magnetic field. As shown in Fig. 12, the *direction of magnetic field lines* is defined to be the direction in which the north end of a compass needle points. The magnetic field is traditionally called the *B-field*.

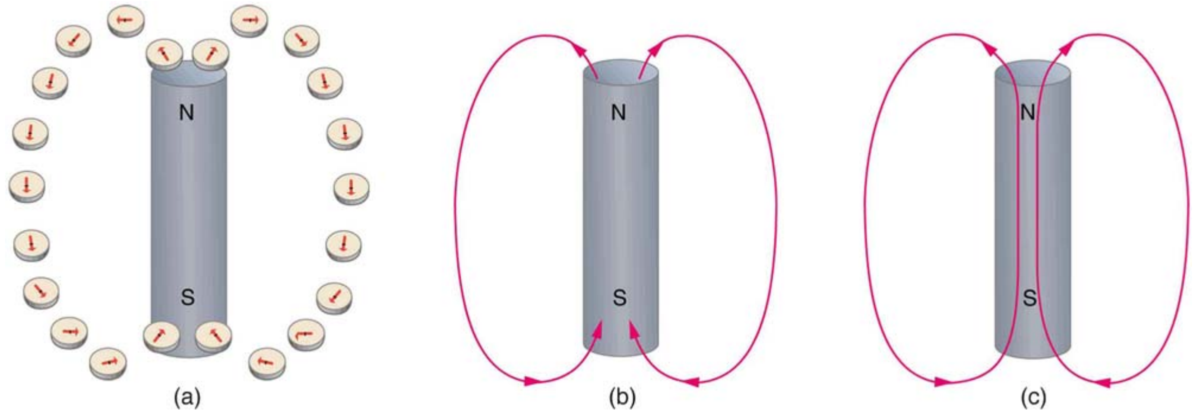

Fig. 12. Magnetic field lines

The magnetic force on a moving charge is one of the most fundamental known. Magnetic force is as important as the electrostatic or Coulomb force. Yet the magnetic force is more complex, in both the number of factors that affects it and in its direction, than the relatively simple Coulomb force. The magnitude of the *magnetic force F* on a charge *q* moving at a speed *v* in a magnetic field of strength *B* is given by

$$F = qvB \sin \theta, \tag{1.3.23}$$

where θ is the angle between the directions of v and B. This force is often called the *Lorentz force*. In fact, this is how we define the magnetic field strength B – in terms of the force on a charged particle moving in a magnetic field. The SI unit for magnetic field strength B is called the *tesla* [T] after the eccentric but brilliant inventor Nikola Tesla (1856–1943). To determine how the tesla relates to other SI units, we solve (1.2.23) for B. The direction of the magnetic force F is perpendicular to the plane formed by v and B, as determined by the *right hand rule 1* (or RHR-1), which is illustrated in Fig. 13. RHR-1 states that, to determine the direction of the magnetic force on a positive moving charge, you point the thumb of the right hand in the direction of v, the fingers in the direction of B, and a perpendicular to the palm points in the direction of F. One way to remember this is that there is one velocity, and so the thumb represents it. There are many field lines, and so the fingers represent them. The force is in the direction you would push with your palm. The force on a negative charge is in exactly the opposite direction to that on a positive charge.

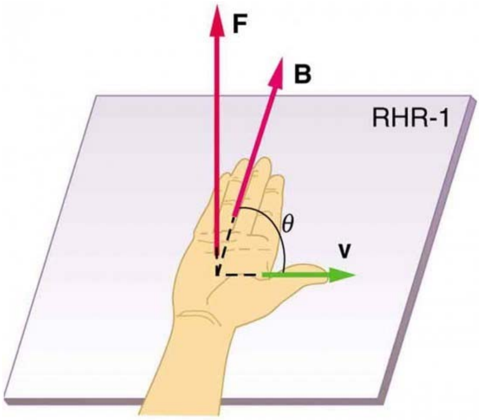

Fig. 13. The direction of the magnetic force on a moving charge is perpendicular to the plane formed by v and B and follows right hand rule 1

Magnetic force can cause a charged particle to move in a circular or spiral path. Cosmic rays are energetic charged particles in outer space, some of which approach the Earth. They can be forced into spiral paths by the Earth's magnetic field. Protons in giant accelerators are kept in a circular path by magnetic force. Because the magnetic force F supplies the centripetal force F_c, we have

$$qvB = \frac{mv^2}{r}, \tag{1.3.24}$$

where r is the radius of curvature of the path of a charged particle with mass m and charge q, moving at a speed v perpendicular to a magnetic field of strength B. If the velocity is not perpendicular to the magnetic field, then v is the component of the velocity perpendicular to the field. The component of the velocity parallel to the field is unaffected, since the magnetic force is zero for motion parallel to the field. This produces a spiral motion rather than a circular one.

We have seen effects of a magnetic field on free-moving charges. The magnetic field also affects charges moving in a conductor. One result is the *Hall effect*, which has important implications and applications. Moving electrons feel a magnetic force toward one side of the conductor, leaving a net positive charge on the other side. This separation of charge creates a voltage Θ, known as *the Hall emf*, across the conductor. The creation of a voltage across a current-carrying conductor by a magnetic field is known as the *Hall effect*, after Edwin Hall, the American physicist who discovered it in 1879. To examine these quantitatively, we need an expression for the Hall emf, Θ, across a conductor. Consider the balance of forces on a moving charge in a situation where B, v, and l are mutually perpendicular. Although the magnetic force moves negative charges to one side, they cannot build up without limit. The electric field caused by their separation opposes the magnetic force, (1.3.23), and the electric force, $F_e = qE$, eventually grows to equal it. That is,

$$qE = qvB \text{ or } E = vB. \tag{1.3.25}$$

Note that the electric field E is uniform across the conductor because the magnetic field B is uniform, as is the conductor. For a uniform electric field, the relationship between electric field and voltage is $E = \Theta/l$, where l is the width of the conductor and ε is the Hall emf. Entering this into the last expression gives

$$\frac{\Theta}{l} = vB, \tag{1.3.26}$$

where Θ is the Hall effect voltage across a conductor of width l through which charges move at a speed v.

Because charges ordinarily cannot escape a conductor, the magnetic force on charges moving in a conductor is transmitted to the conductor itself. We can derive an expression for the magnetic force on a current by taking a sum of the magnetic forces on individual charges. The forces add because they are in the same direction. The force on an individual charge moving at the drift velocity v_d is given by (1.3.23). Taking B to be uniform over a length of wire l and zero elsewhere, the total magnetic force on the wire is then $F = (qv_d B \sin\theta)(N)$, where N is the number of charge carriers in the section of wire of length l. Now, $N = nV$, where n is the number of charge carriers per unit volume and V is the volume of wire in the field. Noting that $V = Al$, where A is the cross-sectional area of the wire, then the force on the wire is $F = (qv_d B \sin\theta)(nAl)$. Gathering terms,

$$F = IlB \sin\theta, \tag{1.3.27}$$

is the equation for magnetic force on a length l of wire carrying a current I in a uniform magnetic field B.

Magnetic fields have both direction and magnitude. As noted before, one way to explore the direction of a magnetic field is with compasses, as shown for a long straight current-carrying wire in Fig. 14. Hall probes can determine the magnitude of the field. The field around a long straight wire is found to be in circular loops. The *right hand rule 2* (RHR-2) emerges from this exploration and is valid for any current segment – point the thumb in the direction of the current, and the fingers curl in the direction of the magnetic field loops created by it.

The magnetic field strength (magnitude) produced by a long straight current-carrying wire is found by experiment to be

$$B = \frac{\mu_0 I}{2\pi r} \text{ (long straight wire)}, \tag{1.3.28}$$

where I is the current, r is the shortest distance to the wire, and the constant $\mu_0 = 4\pi \cdot 10^{-7}$ [T · m/A] is the permeability of free space. (μ_0 is one of the basic constants in nature. Since the wire is very long, the magnitude of the field depends only on distance from the wire r, not on position along the wire.

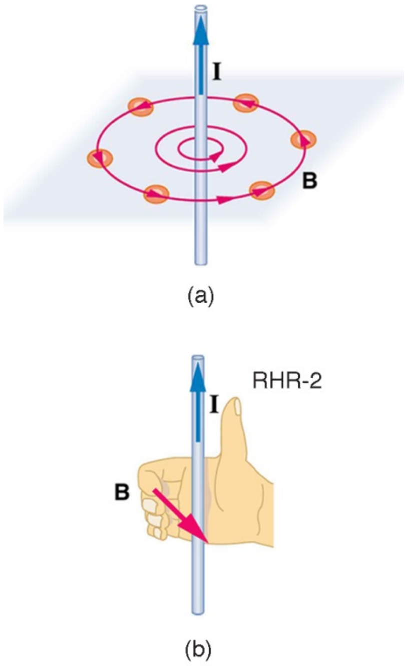

Fig. 14. a – compasses placed near a long straight current-carrying wire indicate that field lines form circular loops centered on the wire; *b* – right hand rule 2 states

A *solenoid* is a long coil of wire (with many turns or loops, as opposed to a flat loop). Because of its shape, the field inside a solenoid can be very uniform, and also very strong. The field just outside the coils is nearly zero. The magnetic field inside of a current-carrying solenoid is very uniform in direction and magnitude. Only near the ends does it begin to weaken and change direction. The field outside has similar complexities to flat loops and bar magnets, but the *magnetic field strength inside a solenoid* is simply

$$B = \mu_0 nI \text{ (inside a solenoid)}, \qquad (1.3.29)$$

where n is the number of loops per unit length of the solenoid $n = N/l$, with N being the number of loops and l the length. Note that B is the field strength anywhere in the uniform region of the interior and not just at the center.

In 1831, the English scientist Michael Faraday (1791–1862) and the American scientist Joseph Henry (1797–1878) independently demonstrated that magnetic fields can produce currents. The basic process of generating emfs (electromotive force) and, hence, currents with magnetic fields is known as *induction*; this process is also called magnetic induction to distinguish it from charging by induction, which utilizes the Coulomb force.

So we see that changing the magnitude or direction of a magnetic field produces an emf. Experiments revealed that there is a crucial quantity called the *magnetic flux*, Φ, given by

$$\Phi = BA\cos\theta, \tag{1.3.30}$$

where B is the magnetic field strength over an area A, at an angle θ with the perpendicular to the area. Any change in magnetic flux Φ induces an emf. This process is defined to be electromagnetic induction. Units of magnetic flux Φ are [T · m^2]. Faraday's experiments showed that the emf induced by a change in magnetic flux depends on only a few factors. First, emf is directly proportional to the change in flux ΔΦ. Second, emf is greatest when the change in time Δt is smallest – that is, emf is inversely proportional to Δt. Finally, if a coil has N turns, an emf will be produced that is N times greater than for a single coil, so that emf is directly proportional to N. The equation for the emf induced by a change in magnetic flux is

$$emp = -N\frac{\Delta\Phi}{\Delta t}. \tag{1.3.31}$$

This relationship is known as Faraday's law of induction. The units for emf are volts, as is usual. The minus sign in Faraday's law of induction is very important. The minus means that the emf creates a current I and magnetic field B that oppose the change in flux ΔΦ – this is known as Lenz's law. The direction (given by the minus sign) of the emf is so important that it is called *Lenz's law* after the Russian Heinrich Lenz (1804–1865), who like Faraday and Henry, independently investigated aspects of induction. Faraday was aware of the direction, but Lenz stated it so clearly that he is credited for its discovery. There are many connections between the electric force and the magnetic force. The fact that a moving electric field produces a magnetic field and, conversely, a moving magnetic field produces an electric field is part of why electric and magnetic forces are now considered to be different manifestations of the same force. This classic unification of electric and magnetic forces into what is called the electromagnetic force is the inspiration for contemporary efforts to unify other basic forces. When the coil of a motor is turned, magnetic flux changes, and an emf (consistent with Faraday's law of induction) is induced. The motor thus acts as a generator whenever its coil rotates. This will happen whether the shaft is turned by an external input, like a belt drive, or by the action of the motor itself. That is, when a motor is doing work and its shaft is turning, an emf is generated. Lenz's law tells us the emf opposes any change, so that the input emf that powers the motor will be opposed by the motor's self-generated emf, called the *back emf* of the motor. Transformers do what their name implies–they transform voltages from one value to another, the term voltage is used rather than emf, because transformers have internal resistance. Induction is the process in which an emf is induced by changing magnetic flux. Many examples have been discussed so far, some more effective than others. Transformers, for example, are designed to be particularly effective at inducing a desired voltage and current with very little loss of energy to other forms. Is there a useful physical quantity related to how "effective" a given device is? The answer is yes, and that physical quantity is called inductance. Mutual inductance is the effect of Faraday's law of induction for one device upon another, such as the primary coil in transmitting energy to the secondary in a transformer. *Self-inductance*, the effect of Faraday's law of induction of a device on itself, also exists. When, for example, current through a coil is increased, the magnetic field and flux also increase, inducing a counter emf, as required by Lenz's law. Conversely, if the current is decreased, an emf is induced that opposes the decrease. Most devices have a fixed geometry, and so the change in flux is due entirely to the change in current ΔI through the device. The induced emf is related to the physical geometry of the device and the rate of change of current. It is given by

$$emp = -L\frac{\Delta I}{\Delta t}, \tag{1.3.32}$$

where L is the self-inductance of the device. A device that exhibits significant self-inductance is called an inductor. The minus sign is an expression of Lenz's law, indicating that emf opposes the change in current. Units of self-inductance are henries [H] just as for mutual inductance. The larger the self-inductance L of a device, the greater its opposition to any change in current through it.

We know from Lenz's law that inductances oppose changes in current. There is an alternative way to look at this opposition that is based on energy. Energy is stored in a magnetic field. It takes time to build up energy, and it also takes time to deplete energy; hence, there is an opposition to rapid change. In an inductor, the magnetic field is directly proportional to current and to the inductance of the device. It can be shown that the *energy stored in an inductor* E_{ind} is given by

$$E_{ind} = \frac{1}{2} L I^2. \qquad (1.3.33)$$

This expression is similar to that for the energy stored in a capacitor.

1.4. Oscillations and Waves

Many systems oscillate, and they have certain characteristics in common. All *oscillations* involve force and energy. Some oscillations create waves. Every *wave* is a disturbance that moves from its source and carries energy. Other examples of waves include earthquakes and visible light. Even subatomic particles, such as electrons, can behave like waves. The time to complete one oscillation remains constant and is called the *period* T. Its units are usually seconds, but may be any convenient unit of time. The word period refers to the time for some event whether repetitive or not; but we shall be primarily interested in periodic motion, which is by definition repetitive. A concept closely related to period is the frequency of an event. *Frequency* f – is defined to be the number of events per unit time. For periodic motion, frequency is the number of oscillations per unit time. The relationship between frequency and period is

$$f = \frac{1}{T}. \qquad (1.4.1)$$

The SI unit for frequency is the cycle per second, which is defined to be a hertz [Hz]: 1 Hz = 1 cycle/sec or $\sec^{-1}$. A cycle is one complete oscillation. Note that a vibration can be a single or multiple events, whereas oscillations are usually repetitive for a significant number of cycles.

The oscillations of a system in which the net force can be described by Hooke's law are of special importance, because they are very common. They are also the simplest oscillatory systems. *Simple Harmonic Motion* (SHM) is the name given to oscillatory motion for a system where the net force can be described by Hooke's law, and such a system is called a simple harmonic oscillator. If the net force can be described by Hooke's law and there is no damping (by friction or other non-conservative forces), then a simple harmonic oscillator will oscillate with equal displacement on either side of the equilibrium position. The maximum displacement from equilibrium is called the *amplitude X*. The units for amplitude and displacement are the same, but depend on the type of oscillation. For the object on the spring, the units of amplitude and displacement are meters; whereas for sound oscillations, they have units of pressure (and other types of oscillations have yet other units). Because amplitude is the maximum displacement, it is related to the energy in the oscillation. The period of a simple harmonic oscillator is given by

$$T = 2\pi\sqrt{\frac{m}{k}}. \tag{1.4.2}$$

In fact, the mass m and the force constant k are the only factors that affect the period and frequency of simple harmonic motion.

The displacement as a function of time t in any simple harmonic motion – that is, one in which the net restoring force can be described by Hooke's law, is given by

$$x(t) = X\cos\frac{2\pi t}{T}, \tag{1.4.3}$$

where X – is amplitude. A $t = 0$, the initial position is $x_0 = X$, and the displacement oscillates back and forth with a period T. (When $t = T$, we get $x = X$ again because $\cos 2\pi = 1$.). Furthermore, from this expression for x, the velocity v as a function of time is given by:

$$v(t) = -v_{\max}\sin\left(\frac{2\pi t}{T}\right), \tag{1.4.4}$$

where $v_{\max} = 2\pi X/T = X\sqrt{k/m}$. The object has zero velocity at maximum displacement – for example, $v = 0$ when $t = 0$, and at that time $x = X$. The minus sign in the first equation for $v(t)$ gives the correct direction for the velocity. Just after the start of the motion, for instance, the velocity is negative because the system is moving back toward the equilibrium point. Finally, we can get an expression for acceleration using Newton's second law. Then we have $x(t)$, $v(t)$, t, and $a(t)$, the quantities needed for kinematics and a description of simple harmonic motion. According to Newton's second law, the acceleration is $a = F/m = kx/m$. So, *a(t)* is also a cosine function:

$$a(t) = -\frac{kX}{m}\cos\frac{2\pi t}{T}. \tag{1.4.5}$$

On the Fig. 15 shows the graphs of *x(t)*, *v(t)*, and *a(t)* versus *t* for the motion of an object on a spring. The net force on the object can be described by Hooke's law, and so the object undergoes simple harmonic motion. Note that the initial position has the vertical displacement at its maximum value *X*; *v* is initially zero and then negative as the object moves down; and the initial acceleration is negative, back toward the equilibrium position and becomes zero at that point. The most important point here is that these equations are mathematically straightforward and are valid for all simple harmonic motion. They are very useful in visualizing waves associated with simple harmonic motion, including visualizing how waves add with one another.

Pendulums are in common usage. A *simple pendulum* is defined to have an object that has a small mass, also known as the pendulum bob, which is suspended from a light wire or string. For the simple pendulum:

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{m}{mg/L}}, \tag{1.4.6}$$

where the force constant is given by $k = mg/L$.

Thus,

$$T = 2\pi\sqrt{\frac{L}{g}} \tag{1.4.7}$$

for the period of a simple pendulum. This result is interesting because of its simplicity. The only things that affect the period of a simple pendulum are its length and the acceleration due to gravity. The period is completely independent of other factors, such as mass.

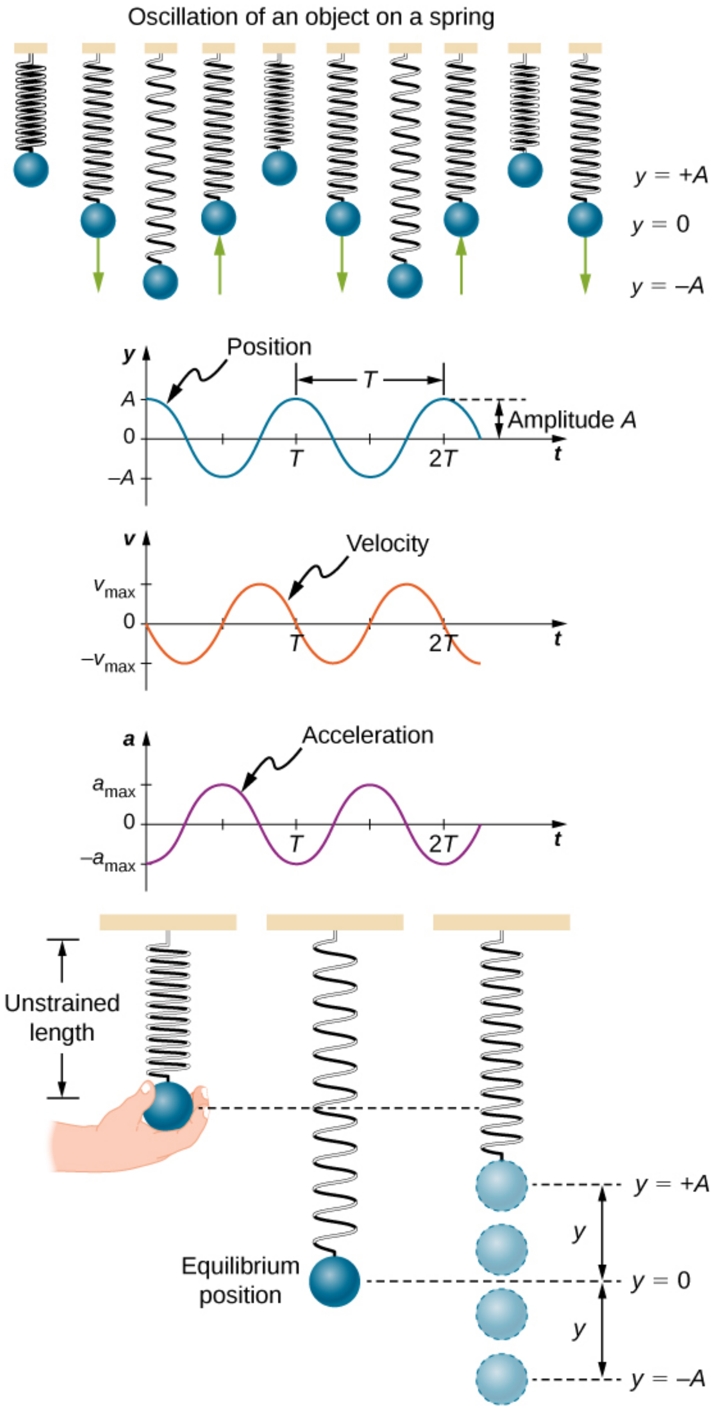

Fig. 15. The simple harmonic motion of an object on a spring and presents graphs of *x(t)*,*v(t)*, and *a(t)* versus time

Because a simple harmonic oscillator has no dissipative forces, the other important form of energy is kinetic energy *KE*. Conservation of energy for these two forms is:

$$KE + PE_{el} = \text{constatnt}. \tag{1.4.8}$$

In the case of undamped simple harmonic motion, the energy oscillates back and forth between kinetic and potential, going completely from one to the other as the system oscillates. So for the simple example of an object on a frictionless surface attached to a spring.

The phenomenon of driving a system with a frequency equal to its natural frequency is called *resonance*. A system being driven at its natural frequency is said to resonate. Wave is a disturbance that propagates, or moves from the place it was created. For water waves, the disturbance is in the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker. For earthquakes, there are several types of disturbances, including disturbance of Earth's surface and pressure disturbances under the surface. Even radio waves are most easily understood using an analogy with water waves. Visualizing water waves is useful because there is more to it than just a mental image. Water waves exhibit characteristics common to all waves, such as amplitude, period, frequency and energy. All wave characteristics can be described by a small set of underlying principles. We define *wave velocity* v_w to be the speed at which the disturbance moves. Wave velocity is sometimes also called the propagation velocity or propagation speed, because the disturbance propagates from one location to another. The water wave also has a length associated with it, called its *wavelength* λ, the distance between adjacent identical parts of a wave (λ is the distance parallel to the direction of propagation). The speed of propagation v_w is the distance the wave travels in a given time, which is one wavelength in the time of one period. In equation form, that is

$$v_w = \frac{\lambda}{T} = f\lambda. \tag{1.4.9}$$

When two or more waves arrive at the same point, they superimpose themselves on one another. More specifically, the disturbances of waves are superimposed when they come together – a phenomenon called superposition. Each disturbance corresponds to a force, and forces add. If the disturbances are along the same line, then the resulting wave is a simple addition of the disturbances of the individual waves – that is, their amplitudes add.

The *Doppler effect* is an alteration in the observed frequency of a sound due to motion of either the source or the observer. For a stationary observer and a moving source, the frequency fobs received by the observer can be shown to be

$$f_{obs} = f_s\left(\frac{v_w}{v_w \pm v_s}\right), \tag{1.4.10}$$

where f_s is the frequency of the source, v_s is the speed of the source along a line joining the source and observer, and v_w is the speed of sound. The minus sign is used for motion toward the observer and the plus sign for motion away from the observer, producing the appropriate shifts up and down in frequency. Note that the greater the speed of the source, the greater the effect. Similarly, for a stationary source and moving observer, the frequency received by the observer fobs is given by

$$f_{obs} = f_s\left(\frac{v_w \pm v_{obs}}{v_w}\right), \tag{1.4.11}$$

where v_{obs} is the speed of the observer along a line joining the source and observer. Here the plus sign is for motion toward the source, and the minus is for motion away from the source.

The energy effects of a wave depend on time as well as amplitude. For example, the longer deep-heat ultrasound is applied, the more energy it transfers. Waves can also be concentrated or spread out. Sunlight, for example, can be focused to burn wood. Earthquakes spread out, so they do less damage the farther they get from the source. In both cases, changing the area the waves cover has important effects. All these pertinent factors are included in the definition of intensity I as power per unit area:

$$I = \frac{P}{A}, \tag{1.4.12}$$

where P is the power carried by the wave through area A. The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter [W/m^2]. For example, infrared and visible energy from the Sun impinge on Earth at an intensity of 1300 W/m^2 just above the atmosphere. There are other intensity-related units in use, too. The most common is the decibel. For example, a 90 decibel sound level corresponds to an intensity of 10−3 W/m^2.

There is a relationship between the *E-* and *B-field* strengths in an electromagnetic wave. This can be understood by again considering the antenna just described. The stronger the *E-field* created by a separation of charge, the greater the current and, hence, the greater the *B-field* created. Since current is directly proportional to voltage (Ohm's law) and voltage is directly proportional to *E-field* strength, the two should be directly proportional. It can be shown that the magnitudes of the fields do have a constant ratio, equal to the speed of light. That is,

$$\frac{E}{B} = c, \tag{1.4.13}$$

where c is the speed of light.

Electromagnetic waves can bring energy into a system by virtue of their *electric and magnetic fields*. These fields can exert forces and move charges in the system and, thus, do work on them. If the frequency of the electromagnetic wave is the same as the natural frequencies of the system (such as microwaves at the resonant frequency of water molecules), the transfer of energy is much more efficient. A wave's energy is proportional to its *amplitude* squared (E^2 or B^2). This is true for waves on guitar strings, for water waves, and for sound waves, where amplitude is proportional to pressure. In electromagnetic waves, the amplitude is the *maximum field strength* of the electric and magnetic fields.

Thus the energy carried and the *intensity* I of an electromagnetic wave is proportional to E^2 and B^2. In fact, for a continuous sinusoidal electromagnetic wave, the average intensity I_{ave} is given by

$$I_{ave} = \frac{E_0 B_0}{2\mu_0}. \tag{1.4.14}$$

The Scotsman James Clerk Maxwell (1831–1879) is regarded as the greatest theoretical physicist of the 19th century. Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by Maxwell's equations, he also developed the kinetic theory of gases and made significant contributions to the understanding of color vision and the nature of Saturn's rings. Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Maxwell's equations are paraphrased here in words because their mathematical statement is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts – why mathematics is the language of science. Maxwell's Theory:

I. Electric field lines originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge, and the strength of the force is related to the electric constant ε_0, also known as the permittivity of free space. From Maxwell's first equation we obtain a special form of Coulomb's law known as Gauss's law for electricity.

II. Magnetic field lines are continuous, having no beginning or end. No magnetic monopoles are known to exist. The strength of the magnetic force is related to the magnetic constant μ_0, also known as the permeability of free space. This second of Maxwell's equations is known as Gauss's law for magnetism.

III. A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. This third of Maxwell's equations is Faraday's law of induction, and includes Lenz's law.

IV. Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell's equations encompasses Ampere's law and adds another source of magnetism – changing electric fields.

There are many types of waves, such as water waves and even earthquakes. Among the many shared attributes of waves are propagation speed, frequency, and wavelength. These are always related by the expression $v_W = f\lambda$. Fig. 16 shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies – that is, it shows the electromagnetic spectrum. Many of the characteristics of the various types of electromagnetic waves are related to their frequencies and wavelengths, as we shall see.

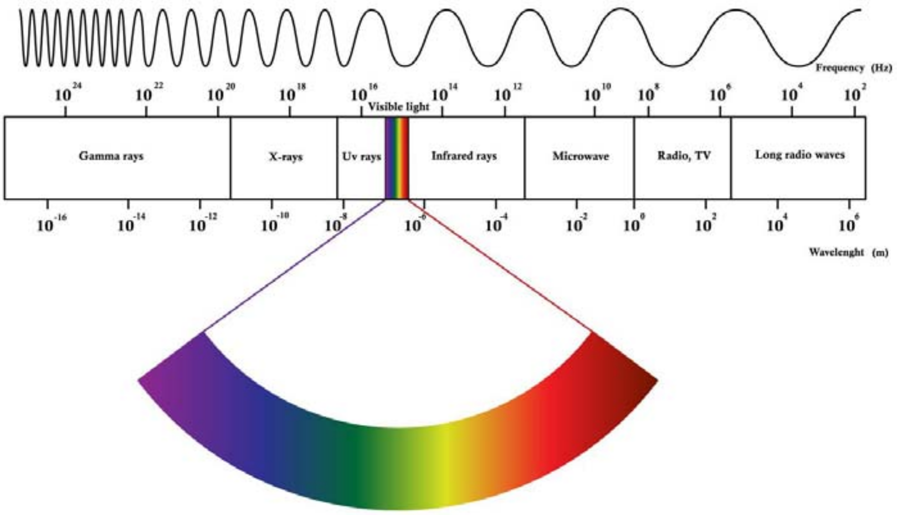

Fig. 16. The electromagnetic spectrum, showing the major categories of electromagnetic waves

Three rules that apply to electromagnetic waves in general are as follows:

I. High-frequency electromagnetic waves are more energetic and are more able to penetrate than low-frequency waves.

II. High-frequency electromagnetic waves can carry more information per unit time than low-frequency waves.

III. The shorter the wavelength of any electromagnetic wave probing a material, the smaller the detail it is possible to resolve.

1.5. Optics

The branch of optics that considers the behavior of light when it exhibits wave characteristics, particularly when it interacts with small objects, is called *wave optics* or sometimes called *physical optics*. The most certain indication of a wave is interference. This wave characteristic is most prominent when the wave interacts with an object that is not large compared with the wavelength. Interference is observed for water waves, sound waves, light waves. Light has wave characteristics in various media as well as in a vacuum. When light goes from a vacuum to some medium, like water, its speed and wavelength change, but its frequency f remains the same. We can think of light as a forced oscillation that must have the frequency of the original source. The speed of light in a medium is $v = c/n$, where n is its index of refraction. If we divide both sides of equation $c = f\lambda$ by n, we get $c/n = f\lambda/n$. This implies that $v = f\lambda_n$, where λ_n is the *wavelength in a medium* and that

$$\lambda_n = \frac{\lambda}{n}, \tag{1.5.1}$$

where λ is the wavelength in vacuum and n is the medium's index of refraction. Therefore, the wavelength of light is smaller in any medium than it is in vacuum. In water, for example, which has $n = 1.333$, the range of visible wavelengths is (380 nm)/1.333 to (760 nm)/1.333 , or $\lambda_n = 285$ to 570 nm. Although wavelengths change while traveling from one medium to another, colors do not, since colors are associated with frequency.

The Dutch scientist Christiaan Huygens (1629–1695) developed a useful technique for determining in detail how and where waves propagate. Starting from some known position, *Huygens's principle* states that: *Every point on a wavefront is a source of wavelets that spread out in the forward direction at the same speed as the wave itself. The new wavefront is a line tangent to all of the wavelets*.

Although Christiaan Huygens thought that light was a wave, Isaac Newton did not. Newton felt that there were other explanations for color, and for the interference and diffraction effects that were observable at the time. Owing to Newton's tremendous stature, his view generally prevailed. The fact that Huygens's principle worked was not considered evidence that was direct enough to prove that light is a wave. The acceptance of the wave character of light came many years later when, in 1801, the English physicist and physician Thomas Young (1773–1829) did his now-classic double slit experiment. Why do we not ordinarily observe wave behavior for light, such as observed in Young's double slit experiment? First, light must interact with something small, such as the closely spaced slits used by Young, to show pronounced wave effects. Furthermore, Young first passed light from a single source (the Sun) through a single slit to make the light somewhat coherent. By coherent, we mean waves are in phase or have a definite phase relationship. Incoherent means the waves have random phase relationships. Why did Young then pass the light through a double slit? The answer to this question is that two slits provide two coherent light sources that then interfere constructively or destructively. Young used sunlight, where each wavelength forms its own pattern, making the effect more difficult to see. We illustrate the double slit experiment with monochromatic (single λ) light to clarify the effect. Fig. 17 shows the pure constructive and destructive interference of two waves having the same wavelength and amplitude.

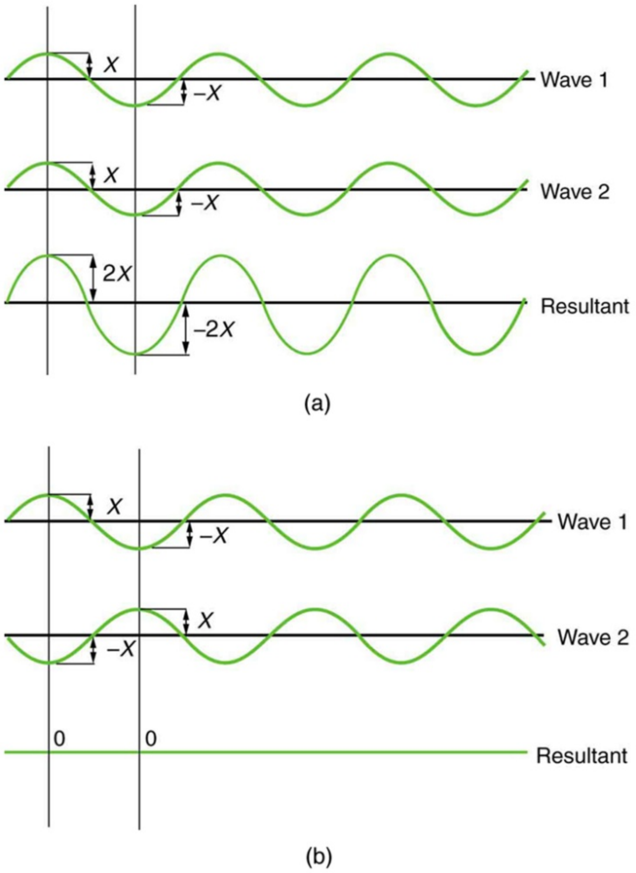

Fig. 17. The amplitudes of waves add: *a* – pure constructive interference is obtained when identical waves are in phase; *b* – pure destructive interference occurs when identical waves are exactly out of phase, or shifted by half a wavelength

When light passes through narrow slits, it is diffracted into semicircular waves, as shown in Fig. 18, *a*. Pure constructive interference occurs where the waves are crest to crest or trough to trough. Pure destructive interference occurs where they are crest to trough. The light must fall on a screen and be scattered into our eyes for us to see the pattern. An analogous pattern for water waves is shown in Fig. 18, *b*. Note that regions of constructive and destructive interference move out from the slits at well-defined angles to the original beam. These angles depend on wavelength and the distance between the slits, as we shall see below.

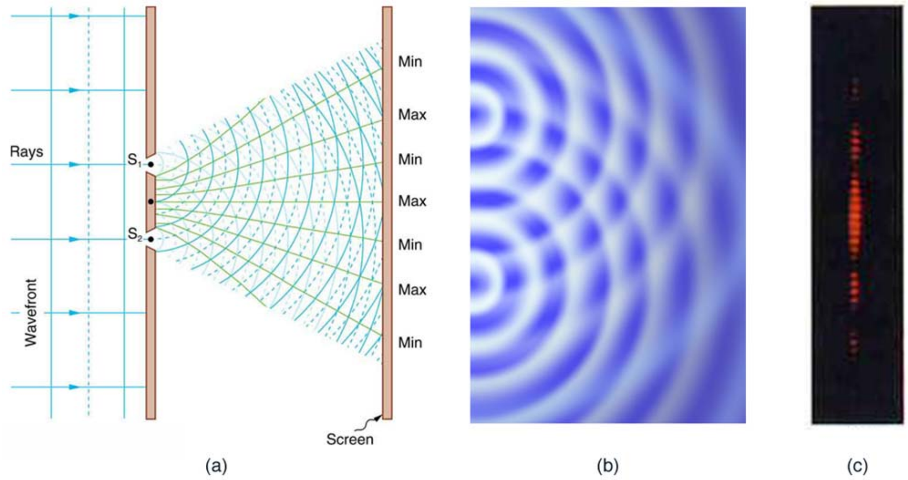

Fig. 18. Double slits produce two coherent sources of waves that interfere

To obtain constructive interference for a double slit, the path length difference must be an integral multiple of the wavelength, or

$$d\sin\theta = m\lambda, \text{for } m = 0, 1, -1, 2, -2, \ldots \text{(constructive)}. \qquad (1.5.2)$$

Similarly, to obtain destructive interference for a double slit, the path length difference must be a half-integral multiple of the wavelength, or

$$d\sin\theta = \left(m + \frac{1}{2}\right)\lambda, \text{for } m = 0, 1, -1, 2, -2, \ldots \text{(distructive)}, \qquad (1.5.3)$$

where λ is the wavelength of the light, *d* is the distance between slits, and θ is the angle from the original direction of the beam as discussed above. We call *m* the *order* of the interference. For example, *m* = 4 is fourth-order interference.

An interesting thing happens if you pass light through a large number of evenly spaced parallel slits, called a *diffraction grating*. An interference pattern is created that is very similar to the one formed by a double slit (see Fig. 19).

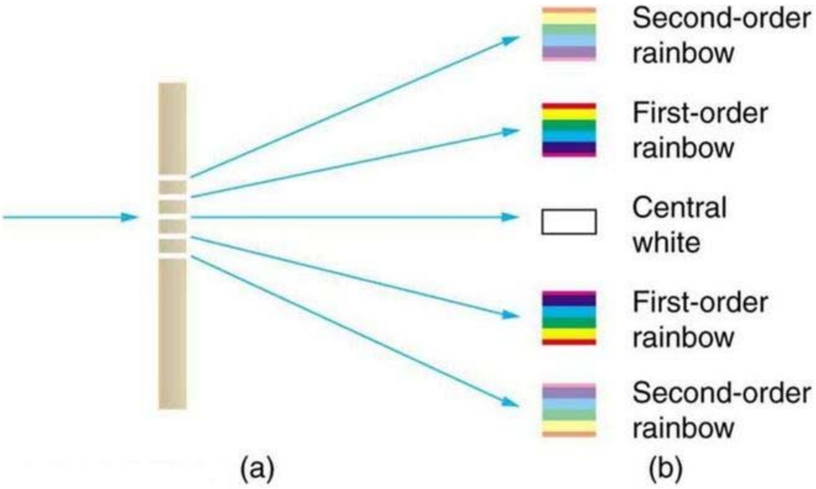

Fig. 19. Diffraction grating and the central maximum is white, and the higher-order maxima disperse white light into a rainbow of colors

The analysis of a diffraction grating is very similar to that for a double slit. As we know from our discussion of double slits in Young's Double Slit Experiment, light is diffracted by each slit and spreads out after passing through. Rays traveling in the same direction (at an angle θ relative to the incident direction) are shown in the figure. Each of these rays travels a different distance to a common point on a screen far away. The rays start in phase, and they can be in or out of phase when they reach a screen, depending on the difference in the path lengths traveled. As seen in the figure, each ray travels a distance $d\sin\theta$ different from that of its neighbor, where d is the distance between slits. If this distance equals an integral number of wavelengths, the rays all arrive in phase, and constructive interference (a maximum) is obtained. Thus, the condition necessary to obtain *constructive interference for a diffraction grating* is

$$d\sin\theta = m\lambda, \text{for } m = 0, 1, -1, 2, -2, \ldots \text{(constructive)}, \qquad (1.5.4)$$

where *d* is the distance between slits in the grating, λ is the wavelength of light, and m is the order of the maximum. Note that this is exactly the same equation as for double slits separated by *d*. However, the slits are usually closer in diffraction gratings than in double slits, producing fewer maxima at larger angles.

To obtain destructive *interference for a single slit*,

$$D\sin\theta = m\lambda, \text{for } m = 1, -1, 2, -2, 3, \ldots \text{(destructive)}, \qquad (1.5.5)$$

where *D* is the slit width, λ is the light's wavelength, θ is the angle relative to the original direction of the light, and m is the order of the minimum.

The *Rayleigh criterion* for the diffraction limit to resolution states that two images are just resolvable when the center of the diffraction pattern of one is directly over the first minimum of the diffraction pattern of the other. See Fig. 20. The first minimum is at an angle of θ, so that two point objects are just resolvable if they are separated by the angle

$$\theta = 1.22\frac{\lambda}{D}, \tag{1.5.6}$$

where λ is the wavelength of light (or other electromagnetic radiation) and *D* is the diameter of the aperture, lens, mirror, etc., with which the two objects are observed. In this expression, θ has units of radians.

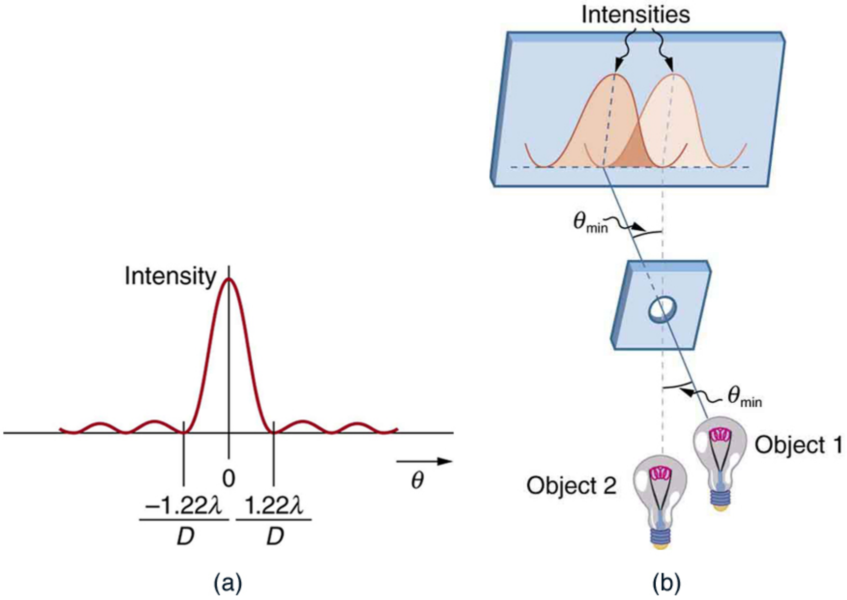

Fig.20. a – graph of intensity of the diffraction pattern for a circular aperture; *b* – two point objects produce overlapping diffraction patterns

Phase change can occur upon reflection. The rule is as follows: *when light reflects from a medium having an index of refraction greater than that of the medium in which it is traveling, a 180^o phase change (or a λ/2 shift) occurs*.

Polarization is the attribute that a wave's oscillations have a definite direction relative to the direction of propagation of the wave. Waves having such a direction are said to be polarized. For an EM wave, we define the direction of polarization to be the direction parallel to the electric field. The oscillations in one rope are in a vertical plane and are said to be *vertically polarized*. Those in the other rope are in a horizontal plane and are *horizontally polarized*. If a vertical slit is placed on the first rope, the waves pass through. However, a vertical slit blocks the horizontally polarized waves. For EM waves, the direction of the electric field is analogous to the disturbances on the ropes. The Sun and many other light sources produce waves that are randomly polarized. Such light is said to be *unpolarized* because it is composed of many waves with all possible directions of polarization. The *axis of a polarizing* filter is the direction along which the filter passes the electric field of an EM wave. Only the component of the EM wave parallel to the axis of a filter is passed. Let us call the angle between the direction of polarization and the axis of a filter θ. If the electric field has an amplitude E, then the transmitted part of the wave has an amplitude $E\cos\theta$ (see Fig. 21). Since the intensity of a wave is proportional to its amplitude squared, the intensity I of the transmitted wave is related to the incident wave by

$$I = I_0\cos^2\theta, \tag{1.5.7}$$

where I_0 is the intensity of the polarized wave before passing through the filter. The above equation is known as Malus's law.

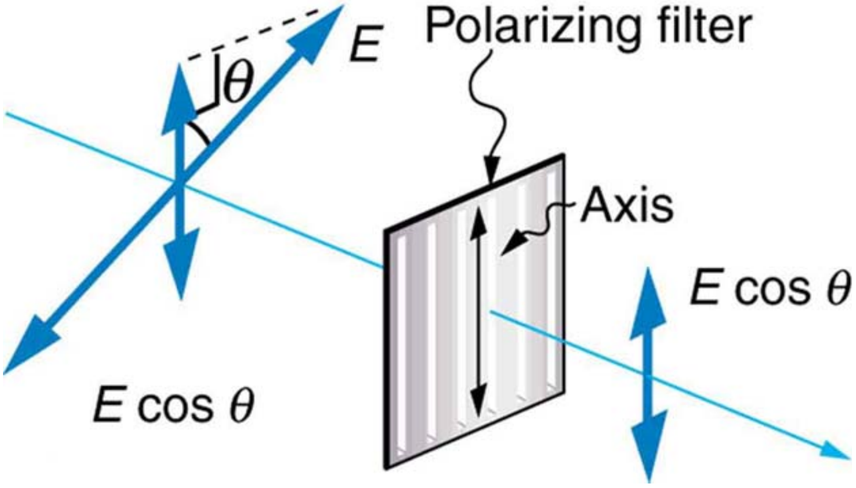

Fig. 21. Polarizing filter transmits only the component of the wave parallel to its axis

Since the part of the light that is not reflected is refracted, the amount of polarization depends on the indices of refraction of the media involved. It can be shown that *reflected light is completely polarized* at an angle of reflection θ_b, given by

$$\tan\theta_b = \frac{n_2}{n_1}, \tag{1.5.8}$$

where n_1 is the medium in which the incident and reflected light travel and n_2 is the index of refraction of the medium that forms the interface that reflects the light. This equation is known as *Brewster's law*, and θ_b is known as *Brewster's angle*, named after the 19th-century Scottish physicist who discovered them.

Many crystals and solutions rotate the plane of polarization of light passing through them. Such substances are said to be *optically active*. Examples include sugar water, insulin, and collagen. In addition to depending on the type of substance, the amount and direction of rotation depends on a number of factors. Among these is the concentration of the substance, the distance the light travels through it, and the wavelength of light. Optical activity is due to the asymmetric shape of molecules in the substance, such as being helical. Measurements of the rotation of polarized light passing through substances can thus be used to measure concentrations, a standard technique for sugars. It can also give information on the shapes of molecules, such as proteins, and factors that affect their shapes, such as temperature and pH. Another interesting phenomenon associated with polarized light is the ability of some crystals to split an unpolarized beam of light into two. Such crystals are said to be *birefringent*. Each of the separated rays has a specific polarization. One behaves normally and is called the ordinary ray, whereas the other does not obey Snell's law and is called the extraordinary ray. Birefringent crystals can be used to produce polarized beams from unpolarized light. Some birefringent materials preferentially absorb one of the polarizations. These materials are called dichroic and can produce polarization by this preferential absorption. This is fundamentally how polarizing filters and other polarizers work.

1.6. Nuclear and Quantum Physics

Quantum mechanics is the branch of physics needed to deal with submicroscopic objects. Because these objects are smaller than we can observe directly with our senses and generally must be observed with the aid of instruments. Quantum mechanics is valid in general, and it must be used rather than classical physics to describe small objects, such as atoms. Atoms, molecules, and fundamental electron and proton charges are all examples of physical entities that are *quantized* – that is, they appear only in certain discrete values and do not have every conceivable value. Quantized is the opposite of continuous. Rather, everything is built of integral multiples of these substructures. Quantum physics is the branch of physics that deals with small objects and the quantization of various entities, including energy and angular momentum. Just as with classical physics, quantum physics has several subfields, such as mechanics and the study of electromagnetic forces. The *correspondence* principle states that in the classical limit (large, slow-moving objects) *quantum mechanics* becomes the same as classical physics. The EM spectrum radiated by a hot solid is linked directly to the solid's temperature. (See Fig. 22.) An ideal radiator is one that has an emissivity of 1 at all wavelengths and, thus, is jet black. Ideal radiators are therefore called *blackbodies*, and their EM radiation is called *blackbody radiation*. Total intensity of the radiation varies as T^4, the fourth power of the absolute temperature of the body, and that the peak of the spectrum shifts to shorter wavelengths at higher temperatures. All of this seems quite continuous, but it was the curve of the spectrum of intensity versus wavelength that gave a clue that the energies of the atoms in the solid are quantized. In fact, providing a theoretical explanation for the experimentally measured shape of the spectrum was a mystery at the turn of the century. When this "ultraviolet catastrophe" was eventually solved, the answers led to new technologies such as computers and the sophisticated imaging techniques.

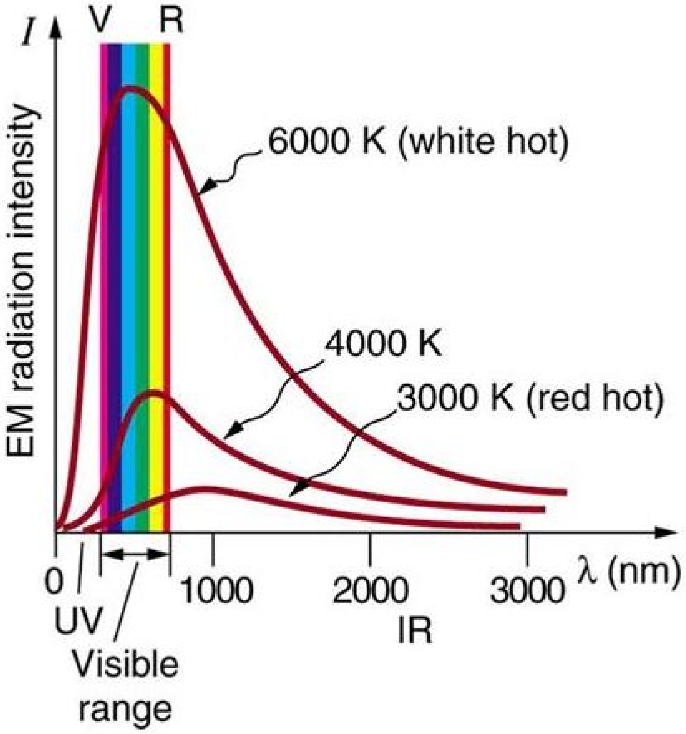

Fig. 22. Graphs of blackbody radiation (from an ideal radiator) at three different radiator temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the peak of the spectrum shifts toward the visible and ultraviolet parts of the spectrum

The German physicist Max Planck (1858–1947) used the idea that atoms and molecules in a body act like oscillators to absorb and emit radiation. The energies of the oscillating atoms and molecules had to be quantized to correctly describe the shape of the blackbody spectrum. Planck deduced that the energy of an oscillator having a frequency f is given by

$$E = \left(n + \frac{1}{2}\right)hf. \tag{1.6.1}$$

Here n is any nonnegative integer (0, 1, 2, 3, …). The symbol h stands for *Planck's constant*, given by

$$h = 6.62607015 \cdot 10^{-34}\ [\mathrm{J \cdot s}]. \tag{1.6.2}$$

Atomic spectra remain an important analytical tool today. Fig. 23 shows an example of an emission spectrum obtained by passing an electric discharge through a material. One of the most important characteristics of these spectra is that they are discrete. By this we mean that only certain wavelengths, and hence frequencies, are emitted. This is called a line spectrum. If frequency and energy are associated as $\Delta E = hf,$ the energies of the electrons in the emitting atoms and molecules are quantized.

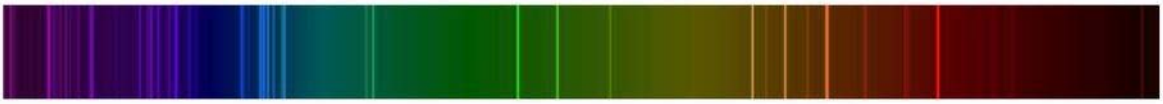

Fig. 23. Emission spectrum of oxygen

When light strikes materials, it can eject electrons from them. This is called the *photoelectric effect*, meaning that light (photo) produces electricity. One common use of the photoelectric effect is in light meters, such as those that adjust the automatic iris on various types of cameras. In a similar way, another use is in solar cells, as you probably have in your calculator or have seen on a roof top or a roadside sign. These make use of the photoelectric effect to convert light into electricity for running different devices. The photoelectric effect has the properties discussed below. All these properties are consistent with the idea that individual photons of EM radiation are absorbed by individual electrons in a material, with the electron gaining the photon's energy. Some of these properties are inconsistent with the idea that EM radiation is a simple wave. For simplicity, let us consider what happens with monochromatic EM radiation in which all photons have the same energy *hf*:

1. If we vary the frequency of the EM radiation falling on a material, we find the following: for a given material, there is a threshold frequency f_0 for the EM radiation below which no electrons are ejected, regardless of intensity. Individual photons interact with individual electrons. Thus if the photon energy is too small to break an electron away, no electrons will be ejected. If EM radiation was a simple wave, sufficient energy could be obtained by increasing the intensity.

2. Once EM radiation falls on a material, electrons are ejected without delay. As soon as an individual photon of a sufficiently high frequency is absorbed by an individual electron, the electron is ejected. If the EM radiation were a simple wave, several minutes would be required for sufficient energy to be deposited to the metal surface to eject an electron.

3. The number of electrons ejected per unit time is proportional to the intensity of the EM radiation and to no other characteristic. High-intensity EM radiation consists of large numbers of photons per unit area, with all photons having the same characteristic energy *hf*.

4. If we vary the intensity of the EM radiation and measure the energy of ejected electrons, we find the following: The maximum kinetic energy of ejected electrons is independent of the intensity of the EM radiation. Since there are so many electrons in a material, it is extremely unlikely that two photons will interact with the same electron at the same time, thereby increasing the energy given it. Instead (as noted in 3 above), increased intensity results in more electrons of the same energy being ejected. If EM radiation were a simple wave, a higher intensity could give more energy, and higher-energy electrons would be ejected.

5. The kinetic energy of an ejected electron equals the photon energy minus the binding energy of the electron in the specific material. An individual photon can give all of its energy to an electron. The photon's energy is partly used to break the electron away from the material. The remainder goes into the ejected electron's kinetic energy. In equation form, this is given by

$$KE_e = hf - BE, \tag{1.6.3}$$

where KE_e is the maximum kinetic energy of the ejected electron, hf is the photon's energy, and BE is the *binding energy* of the electron to the particular material. (BE is sometimes called the *work function* of the material.) This equation, due to Einstein in 1905, explains the properties of the photoelectric effect quantitatively. An individual photon of EM radiation (it does not come any other way) interacts with an individual electron, supplying enough energy, BE, to break it away, with the remainder going to kinetic energy. The binding energy is $BE = hf_{0,}$ where f_0 is the threshold frequency for the particular material.

A photon is a quantum of EM radiation. Its energy is given by $E = hf$ and is related to the frequency f and wavelength λ of the radiation by

$$E = hf = \frac{hc}{\lambda} \text{ (energy of a photon),} \tag{1.6.4}$$

where E is the energy of a single photon and c is the speed of light.

All EM radiation is composed of photons. Photons act as individual quanta and interact with individual electrons, atoms, molecules, and so on. The energy a photon carries is, thus, crucial to the effects it has. Table 8 lists representative submicroscopic energies in eV.

Table 8

Representative Energies for Submicroscopic Effects

Rotational energies of molecules	10^{-5} eV
Vibrational energies of molecules	0.1 eV
Energy between outer electron shells in atoms	1 eV
Binding energy of a weakly bound molecule	1 eV
Energy of red light	2 eV
Binding energy of a tightly bound molecule	10 eV
Energy to ionize atom or molecule	10 to 1000 eV

The quantum of EM radiation we call a *photon* has properties analogous to those of particles we can see, such as grains of sand. A photon interacts as a unit in collisions or when absorbed, rather than as an extensive wave. Massive quanta, like electrons, also act like macroscopic particles – something we expect, because they are the smallest units of matter. Particles carry momentum as well as energy. Despite photons having no mass, there has long been evidence that EM radiation carries momentum. (Maxwell and others who studied EM waves predicted that they would carry momentum.) It is now a well-established fact that photons do have momentum. In fact, photon momentum is suggested by the photoelectric effect, where photons knock electrons out of a substance.

Not only is momentum conserved in all realms of physics, but all types of particles are found to have momentum. We expect particles with mass to have momentum, but now we see that massless particles including photons also carry momentum. Momentum is conserved in quantum mechanics just as it is in relativity and classical physics. Some of the earliest direct experimental evidence of this came from scattering of x-ray photons by electrons in substances, named Compton scattering after the American physicist Arthur H. Compton (1892–1962). Around 1923, Compton observed that x rays scattered from materials had a decreased energy and correctly analyzed this as being due to the scattering of photons from electrons. This phenomenon could be handled as a collision between two particles – a photon and an electron at rest in the material. Energy and momentum are conserved in the collision. He won a Nobel Prize in 1929 for the discovery of this scattering, now called the *Compton Effect*, because it helped prove that *photon momentum* is given by

$$p = \frac{h}{\lambda}, \tag{1.6.5}$$

where h is Planck's constant and λ is the photon wavelength. We can see that photon momentum is small, since $p = h/\lambda$ and h is very small. It is for this reason that we do not ordinarily observe photon momentum. Our mirrors do not recoil when light reflects from them. Compton saw the effects of photon momentum because he was observing x rays, which have a small wavelength and a relatively large momentum, interacting with the lightest of particles, the electron. In 1923 a French physics graduate student named Prince Louis-Victor de Broglie (1892–1987) made a radical proposal based on the hope that nature is symmetric. If EM radiation has both particle and wave properties, then nature would be symmetric if matter also had both particle and wave properties. If what we once thought of as an unequivocal wave (EM radiation) is also a particle, then what we think of as an unequivocal particle (matter) may also be a wave. De Broglie's suggestion, made as part of his doctoral thesis, was so radical that it was greeted with some skepticism. A copy of his thesis was sent to Einstein, who said it was not only probably correct, but that it might be of fundamental importance. With the support of Einstein and a few other prominent physicists, de Broglie was awarded his doctorate. De Broglie took both relativity and quantum mechanics into account to develop the proposal that all particles have a wavelength, given by

$$\lambda = \frac{h}{p} \text{ (matter and photons)}, \tag{1.6.6}$$

where h is Planck's constant and p is momentum. This is defined to be the *de Broglie wavelength*. Note that we already have this for photons, from the equation (1.6.5). The hallmark of a wave is interference. All microscopic particles, whether massless, like photons, or having mass, like electrons, have wave properties. The relationship between momentum and wavelength is fundamental for all particles.

After de Broglie proposed the wave nature of matter, many physicists, including Schrödinger and Heisenberg, explored the consequences. The idea quickly emerged that, because of its wave character, a particle's trajectory and destination cannot be precisely predicted for each particle individually. However, each particle goes to a definite place (as illustrated in Fig. 24). After compiling enough data, you get a distribution related to the particle's wavelength and diffraction pattern. There is a certain probability of finding the particle at a given location, and the overall pattern is called a *probability distribution*. Those who developed quantum mechanics devised equations that predicted the probability distribution in various circumstances.

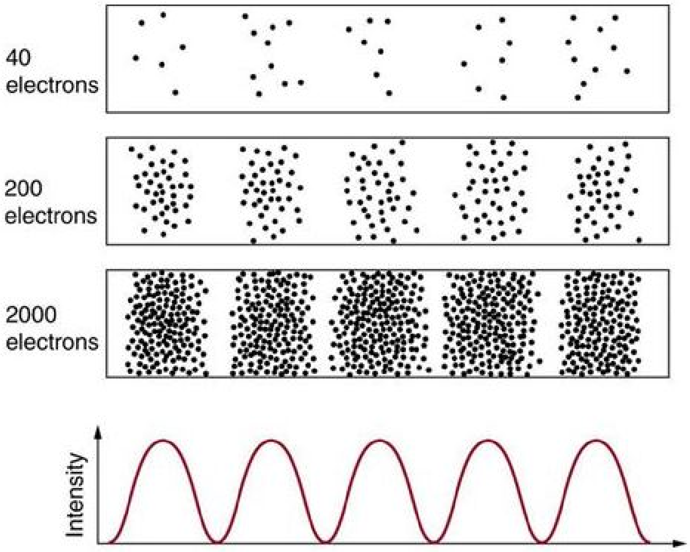

Fig. 24. The building up of the diffraction pattern of electrons scattered from a crystal surface

It was Werner Heisenberg who first stated the limit to knowledge in 1929 as a result of his work on quantum mechanics and the wave characteristics of all particles. Specifically, consider simultaneously measuring the position and momentum of an electron (it could be any particle). There is an *uncertainty in position* Δx that is approximately equal to the wavelength of the particle. That is,

$$\Delta x \approx \lambda. \tag{1.6.7}$$

If the electron's position is measured repeatedly, a spread in locations will be observed, implying an uncertainty in position Δx. To detect the position of the particle, we must interact with it, such as having it collide with a detector. In the collision, the particle will lose momentum. This change in momentum could be anywhere from close to zero to the total momentum of the particle, $p = h/\lambda$. It is not possible to tell how much momentum will be transferred to a detector, and so there is an *uncertainty in momentum* Δp, too. In fact, the uncertainty in momentum may be as large as the momentum itself, which in equation form means that

$$\Delta p =\approx \frac{h}{\lambda}. \tag{1.6.8}$$

The uncertainty in position can be reduced by using a shorter-wavelength electron, since (1.6.7). But shortening the wavelength increases the uncertainty in momentum, since (1.6.8). Conversely, the uncertainty in momentum can be reduced by using a longer-wavelength electron, but this increases the uncertainty in position. Mathematically, you can express this trade-off by multiplying the uncertainties. The wavelength cancels, leaving

$$\Delta xp \approx h. \tag{1.6.9}$$

So if one uncertainty is reduced, the other must increase so that their product is $\approx h$. With the use of advanced mathematics, Heisenberg showed that the best that can be done in a simultaneous measurement of position and momentum is

$$\Delta x \Delta p \geq \frac{h}{4\pi}. \tag{1.6.10}$$

This is known as the *Heisenberg uncertainty principle*. It is impossible to measure position x and momentum p simultaneously with uncertainties Δx and Δp that multiply to be less than $h/4\pi$. Neither uncertainty can be zero. Neither uncertainty can become small without the other becoming large. A small wavelength allows accurate position measurement, but it increases the momentum of the probe to the point that it further disturbs the momentum of a system being measured. For example, if an electron is scattered from an atom and has a wavelength small enough to detect the position of electrons in the atom, its momentum can knock the electrons from their orbits in a manner that loses information about their original motion. It is therefore impossible to follow an electron in its orbit around an atom. If you measure the electron's position, you will find it in a definite location, but the atom will be disrupted. Repeated measurements on identical atoms will produce interesting probability distributions for electrons around the atom, but they will not produce motion information. The probability distributions are referred to as electron clouds or orbitals. There is another form of *Heisenberg's uncertainty principle for simultaneous measurements of energy and time*. In equation form,

$$\Delta E \Delta t \geq \frac{h}{4\pi}, \tag{1.6.11}$$

where ΔE is the *uncertainty in energy* and Δt is the *uncertainty in time*. This means that within a time interval Δt, it is not possible to measure energy precisely – there will be an uncertainty ΔE in the measurement. In order to measure energy more precisely (to make ΔE smaller), we must increase Δt. This time interval may be the amount of time we take to make the measurement, or it could be the amount of time a particular state exists.

People have long speculated about the structure of matter and the existence of atoms. The earliest significant ideas to survive are due to the ancient Greeks in the fifth century BCE, especially those of the philosophers Leucippus and Democritus. (There is some evidence that philosophers in both India and China made similar speculations, at about the same time.) They considered the question of whether a substance can be divided without limit into ever smaller pieces. There are only a few possible answers to this question. One is that infinitesimally small subdivision is possible. Another is what Democritus in particular believed –

that there is a smallest unit that cannot be further subdivided. Democritus called this the *atom*. We now know that atoms themselves can be subdivided, but their identity is destroyed in the process, so the Greeks were correct in a respect. The Greeks also felt that atoms were in constant motion, another correct notion. Knowledge of the properties of elements and compounds grew, culminating in the mid-19th-century development of the periodic table of the elements by Dmitri Mendeleev (1834–1907), the great Russian chemist. Mendeleev proposed an ingenious array that highlighted the periodic nature of the properties of elements. Believing in the systematics of the periodic table, he also predicted the existence of then-unknown elements to complete it. Once these elements were discovered and determined to have properties predicted by Mendeleev, his periodic table became universally accepted. The first truly direct evidence of atoms is credited to Robert Brown, a Scottish botanist. In 1827, he noticed that tiny pollen grains suspended in still water moved about in complex paths. This can be observed with a microscope for any small particles in a fluid. The motion is caused by the random thermal motions of fluid molecules colliding with particles in the fluid, and it is now called *Brownian motion*.

It was Albert Einstein who, starting in his epochal year of 1905, published several papers that explained precisely how Brownian motion could be used to measure the size of atoms and molecules. (In 1905 Einstein created special relativity, proposed photons as quanta of EM radiation, and produced a theory of Brownian motion that allowed the size of atoms to be determined. All of this was done in his spare time, since he worked days as a patent examiner. Any one of these very basic works could have been the crowning achievement of an entire career – yet Einstein did even more in later years.) Their sizes were only approximately known to be 10^{-10} m, based on a comparison of latent heat of vaporization and surface tension made in about 1805 by Thomas Young of double-slit fame and the famous astronomer and mathematician Simon Laplace. Using Einstein's ideas, the French physicist Jean-Baptiste Perrin (1870–1942) carefully observed Brownian motion; not only did he confirm Einstein's theory, he also produced accurate sizes for atoms and molecules. Since molecular weights and densities of materials were well established, knowing atomic and molecular sizes allowed a precise value for Avogadro's number to be obtained. (If we know how big an atom is, we know how many fit into a certain volume.) Perrin also used these ideas to explain atomic and molecular agitation effects in sedimentation, and he received the 1926 Nobel Prize for his achievements. Most scientists were already convinced of the existence of atoms, but the accurate observation and analysis of Brownian motion was conclusive – it was the first truly direct evidence.

The English physicist J. J. Thomson (1856–1940) improved and expanded the scope of experiments with gas discharge tubes. He verified the negative charge of the cathode rays with both magnetic and electric fields. Additionally, he collected the rays in a metal cup and found an excess of negative charge. Thomson was also able to measure the ratio of the charge of the electron to its mass, q_e/m_e – an important step to finding the actual values of both q_e and m_e. Thomson performed a variety of experiments using differing gases in discharge tubes and employing other methods, such as the photoelectric effect, for freeing electrons from atoms. He always found the same properties for the electron, proving it to be an independent particle. Thomson was awarded the 1906 Nobel Prize in Physics. In retrospect, it is difficult to appreciate how astonishing it was to find that the atom has a substructure. An American physicist, Robert Millikan (1868–1953), decided to improve upon Thomson's experiment for measuring q_e and was eventually forced to try another approach, which is now a classic experiment performed by students. With the charge of the electron known and the charge-to-mass ratio known, the electron's mass can be calculated. It is

$$m = \frac{q_e}{\left(\frac{q_e}{m_e}\right)}, \tag{1.6.12}$$

or $m_e = 9.10938215 \cdot 10^{-31}$ [kg].

A similar calculation gives the masses of other particles, including the proton $m_p = 1.67262192369 \cdot 10^{-27}$ [kg].

Based on the size and mass of the nucleus revealed by his experiment, as well as the mass of electrons, Rutherford proposed the *planetary model of the atom*. The planetary model of the atom pictures low-mass electrons orbiting a large-mass nucleus. The sizes of the electron orbits are large compared with the size of the nucleus, with mostly vacuum inside the atom. This picture is analogous to how low-mass planets in our solar system orbit the large-mass Sun at distances large compared with the size of the sun. In the atom, the attractive Coulomb force is analogous to gravitation in the planetary system. (See Fig. 25).

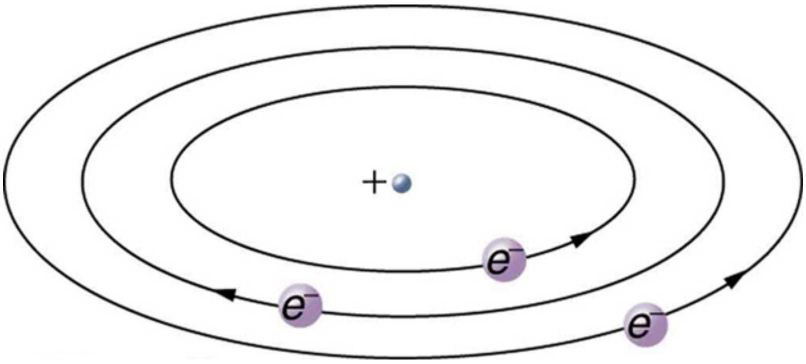

Fig. 25. Rutherford’s planetary model of the atom incorporates the characteristics of the nucleus, electrons, and the size of the atom

The great Danish physicist Niels Bohr (1885–1962) made immediate use of Rutherford’s planetary model of the atom. Bohr became convinced of its validity and spent part of 1912 at Rutherford’s laboratory. In 1913, after returning to Copenhagen, he began publishing his theory of the simplest atom, hydrogen, based on the planetary model of the atom. For decades, many questions had been asked about atomic characteristics. From their sizes to their spectra, much was known about atoms, but little had been explained in terms of the laws of physics. Bohr’s theory explained the atomic spectrum of hydrogen and established new and broadly applicable principles in quantum mechanics. Bohr was able to derive the formula for the hydrogen spectrum using basic physics, the planetary model of the atom, and some very important new proposals. His first proposal is that only certain orbits are allowed: we say that the orbits of electrons in atoms are quantized. Each orbit has a different energy, and electrons can move to a higher orbit by absorbing energy and drop to a lower orbit by emitting energy. If the orbits are quantized, the amount of energy absorbed or emitted is also quantized, producing discrete spectra. Photon absorption and emission are among the primary methods of transferring energy into and out of atoms. The energies of the photons are quantized, and their energy is explained as being equal to the change in energy of the electron when it moves from one orbit to another. In equation form, this is

$$\Delta E = hf = E_i - E_f. \tag{1.6.13}$$

Here, ΔE is the change in energy between the initial and final orbits, and hf is the energy of the absorbed or emitted photon. It is quite logical (that is, expected from our everyday experience) that energy is involved in changing orbits. Fig. 26 shows an *energy-level diagram*, a convenient way to display energy states. In the present discussion, we take these to be the allowed energy levels of the electron. Energy is plotted vertically with the lowest or ground state at the bottom and with excited states above. Given the energies of the lines in an atomic spectrum, it is possible (although sometimes very difficult) to determine the energy levels of an atom. Energy-level diagrams are used for many systems, including molecules and nuclei. A theory of the atom or any other system must predict its energies based on the physics of the system.

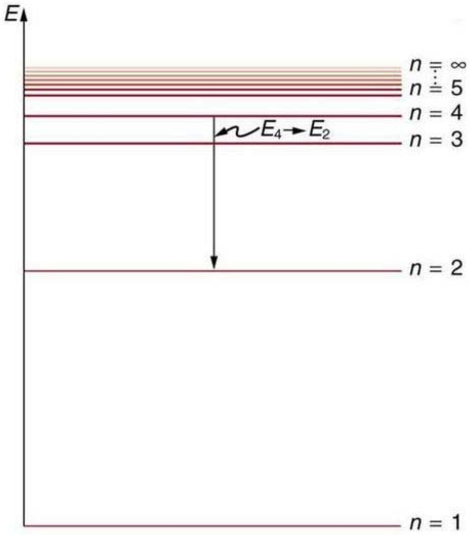

Fig. 26. An energy-level diagram plots energy vertically and is useful in visualizing the energy states of a system and the transitions between them

Assuming circular orbits, Bohr proposed that the *angular momentum* L of an electron in its orbit is quantized, that is, it has only specific, discrete values. The value for L is given by the formula

$$L = m_e v r_n = n\frac{h}{2\pi}\ (n = 1, 2, 3, \ldots), \qquad (1.6.14)$$

where L is the angular momentum, m_e is the electron's mass, r_n is the radius of the n th orbit, and h is Planck's constant.

Angular momentum quantization is stated in an earlier equation. We solve that equation for v, substitute it into the above, and rearrange the expression to obtain the radius of the orbit. This yields:

$$r_n = \frac{n^2}{Z} a_B, \text{for allowed orbits (n} = 1, 2, 3, \ldots), \qquad (1.6.15)$$

where a_B is defined to be the *Bohr radius*, since for the lowest orbit ($n = 1$) and for hydrogen ($Z = 1$) , $r_1 = a_B$. These last two equations can be used to calculate the radius of the allowed (quantized) electron orbits in any hydrogen-like atom. Now we substitute r_n and v from earlier equations into the above expression for energy. Algebraic manipulation yields

$$E_n = -\frac{Z^2}{n^2} E_0 \ (n = 1, 2, 3, \ldots), \qquad (1.6.16)$$

for the orbital energies of hydrogen-like atoms. Here, E_0 is the ground-state energy ($n = 1$) for hydrogen ($Z = 1$). Electron orbital energies are quantized in all atoms and molecules. Angular momentum is quantized. The electrons do not spiral into the nucleus, as expected classically, accelerated charges radiate, so that the electron orbits classically would decay quickly, and the electrons would sit on the nucleus – matter would collapse.

Physical characteristics that are quantized – such as energy, charge, and angular momentum – are of such importance that names and symbols are given to them. The values of quantized entities are expressed in terms of quantum numbers, and the rules governing them are of the utmost importance in determining what nature is and does. The fact that the magnitude of angular momentum is quantized was first recognized by Bohr in relation to the hydrogen atom; it is now known to be true in general. With the development of quantum mechanics, it was found that the magnitude of angular momentum L can have only the values

$$L = \sqrt{l(l+1)} \frac{h}{2\pi} \ (l = 0, 1, 2, \ldots, n-1), \qquad (1.6.17)$$

where l is defined to be the *angular momentum quantum number*. The rule for l in atoms is given in the parentheses. Intrinsic angular momentum is quantized independently of orbital angular momentum. Additionally, the direction of the spin is also quantized. It has been found that the *magnitude of the intrinsic (internal) spin angular momentum, S*, of an electron is given by

$$S = \sqrt{s(s+1)}\frac{h}{2\pi}\left(s = \frac{1}{2}\text{for electrons}\right), \tag{1.6.18}$$

where s is defined to be the spin quantum number.

In 1925, the Austrian physicist Wolfgang Pauli proposed the following rule: No two electrons can have the same set of quantum numbers. That is, no two electrons can be in the same state. This statement is known as the *Pauli exclusion principle*, because it excludes electrons from being in the same state. The Pauli exclusion principle is extremely powerful and very broadly applicable. It applies to any identical particles with half-integral intrinsic spin – hat is, having s = 1/2, 3/2, ... Thus no two electrons can have the same set of quantum numbers.

In 1896, the French physicist Antoine Henri Becquerel (1852–1908) accidentally found that a uranium-rich mineral called pitchblende emits invisible, penetrating rays that can darken a photographic plate enclosed in an opaque envelope. The rays therefore carry energy; but amazingly, the pitchblende emits them continuously without any energy input. This is an apparent violation of the law of conservation of energy, one that we now understand is due to the conversion of a small amount of mass into energy, as related in Einstein's famous equation $E = mc^2$. It was soon evident that Becquerel's rays originate in the nuclei of the atoms and have other unique characteristics. The emission of these rays is called *nuclear radioactivity* or simply *radioactivity*. The rays themselves are called *nuclear radiation*. A nucleus that spontaneously destroys part of its mass to emit radiation is said to *decay* (a term also used to describe the emission of radiation by atoms in excited states). A substance or object that emits nuclear radiation is said to be *radioactive*. Research begun by people such as New Zealander Ernest Rutherford soon after the discovery of nuclear radiation indicated that different types of rays are emitted. Eventually, three types were distinguished and named *alpha* (α), *beta* (β), and *gamma* (γ), because, like x-rays, their identities were initially unknown. The range of radiation is defined to be the distance it can travel through a material. Range is related to several factors, including the energy of the radiation, the material encountered, and the type of radiation.

All properties of a nucleus are determined by the number of protons and neutrons it has. A specific combination of protons and neutrons is called a *nuclide* and is a unique nucleus. The following notation is used to represent a particular nuclide:

$$ {}_{Z}^{A}X_{N}, \tag{6.19}$$

where the symbols A, X, Z, and N are defined as follows: The number of protons in a nucleus is the atomic number Z. The symbol A is defined as the number of nucleons or the total number of protons and neutrons,

$$A = N + Z, \tag{1.6.20}$$

where A is also called the mass number. This name for A is logical; the mass of an atom is nearly equal to the mass of its nucleus, since electrons have so little mass.

There is a tremendous range in the half-lives of various nuclides, from as short as 10^{-23} [s] for the most unstable, to more than 10^{16} [yr] for the least unstable, or about 46 orders of magnitude. Nuclides with the shortest half-lives are those for which the nuclear forces are least attractive, an indication of the extent to which the nuclear force can depend on the particular combination of neutrons and protons. The concept of half-life is applicable to other subatomic particles. It is also applicable to the decay of excited states in atoms and nuclei. The following equation gives the quantitative relationship between the original number of nuclei present at time zero (N_0) and the number (N) at a later time t:

$$N = N_0 e^{-\lambda t}, \tag{1.6.21}$$

where $e = 2.71828...$ is the base of the natural logarithm, and λ is the decay constant for the nuclide. The shorter the half-life, the larger is the value of λ, and the faster the exponential $e^{-\lambda t}$ decreases with time.

2. LABORATORY PRACTICE

2.1. Introduction to statistics, error and measurement

When you do an experiment, it is important to be able to evaluate how well you can trust your measurements. For example, the known value of *g*, the acceleration due to gravity, is ≈ 9.81 [m · s^{-2}], ("≈" means approximately equal to). If you make a measurement that says $g = 10.1$ [m · s^{-2}], is that measurement "wrong"? How do you compare that measurement to the known value of *g*? Suppose you measure some quantity that is not known? You may make a number of measurements, and get several different results. For example, suppose you measure the mass of an object three times, and get three different values, 5 [kg], 4.8 [kg], and 5.4 [kg]. Can you evaluate what the real mass of the object is from those measurements?

The mathematical tools we will learn in this lab will answer some of these questions. They are some of the most basic methods of statistical analysis; they will allow us to give information about our measurements in a standard, concise way, and to evaluate how "correct" our measurements are. The methods we will cover are used in all areas of science which involve taking any measurements, from popularity polls of politicians, to evaluating the results of a clinical trial, to making precise measurements of basic physical quantities. Let's start with the basics of the different kinds of *errors*, and how to measure them.

There are two types of errors encountered in experimental physics: *systematic errors* and *random errors*.

Systematic errors can be introduced

- by the design of the experiment;
- by problems with the instruments you are using to take your data;
- by your own biases.

Consider a very simple experiment designed to measure the dimensions of a particular piece of material precisely. A systematic error of could be introduced if the measuring instrument is calibrated improperly. For example, a scale might be set a little too low, so that what reads as "zero" is really "–1 kg". Everything you measure on the scale will come out one kilogram lighter than it really is. If a particular observer always tends to overestimate the size of a measurement, that would also be a systematic error, but one related to the personal characteristics of the experimenter.

Random errors are produced by unpredictable and uncontrollable variations in the experiment. These can be due to the limits of the precision of the measuring device, or due to the experimenter's inability to make the same measurement in precisely the same way each time. Even if systematic errors can be eliminated by good experimental design, there will always be some uncertainty due to random errors. Numerical values measured in experiments are therefore never absolutely precise; there is always some uncertainty.

The *accuracy of a measurement* describes how close the experimental result comes to the actual value. That is, it is a measure of the "correctness" of the result. For example, if two independent experiments give the values 2.717 and 2.659 for e (the base of the natural log), the first value is said to be more accurate because the actual value of e is 2.718… The *precision of an experiment* is a measure of the reproducibility of the result. Suppose you measure the same thing three times. The precision would be a measure of how similar all the measurements are to each other. It is a measure of the magnitude of uncertainty in the result. The object of some experiments is to measure the value of a well-known quantity, such as g. (You'll be making this measurement yourselves in next week's experiment!) The most accurate value of these quantities (measured by teams of dedicated professional scientists!!) is the value given in your textbooks and tables. In making a comparison between the results of your experiment and the accepted value measured with much more precision in specialized laboratories, you want to cite the *percent error*, a measurement of how much your measurement differs from the "official" value. The *absolute difference* between the experimental value E and the accepted value A is written |E-A|, where the "|" signs mean absolute value. The *fractional error* is the ratio of this *absolute difference* over the accepted value:

$$\text{Fractional error} = \frac{|E-A|}{A}. \tag{2.1.1}$$

Usually, convert this fractional error into a percent, and give the *percent error*:

$$\text{Percent error} = \frac{|E-A|}{A} \times 100\,\%. \tag{2.1.2}$$

Even if systematic errors can be eliminated from an experiment, the *mean value of a set of measurements of a quantity*, *x*, is a better estimate of the true value of x than is any single measurement. For this reason, experiments are often repeated a number of times. If we denote <*x*> as the mean value, and there are *N* measurements x_i (where *i* varies from 1 to *N*), then <*x*> is defined by the following equation:

$$\langle x \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{2.1.3}$$

To obtain the *mean deviation* of a set of *N* measurements, the absolute deviations of $|\Delta x_\mathrm{i}|$ are determined; that is $|\varDelta x_i| = |x_i - \langle x \rangle|$.

The *mean deviation* <Δ*x*> is then

$$\langle \Delta x \rangle = \frac{1}{N}\sum_{i=1}^{N} |\Delta x_i|. \tag{2.1.4}$$

Usually, the report an experimental measurement as the mean "plus or minus" the mean deviation: $E = \langle x \rangle \pm \langle \varDelta x \rangle$.

Statistical theory states that the precision of a measurement can be determined using a quantity called the *standard deviation*, *σ* (called "sigma", this is the Greek lower-case "s"). The standard deviation of a distribution of measurements is defined as follows:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \langle x \rangle)^2}. \tag{2.1.5}$$

The standard deviation is a measure of spread. If the standard deviation is small, then the spread in the measured values about the mean is small, and so the precision in the measurements is high. The standard deviation is always positive and has the same units as the measured values. It can be shown, for a *Gaussian distribution*, that 69 % of the data points will fall within one standard deviation of <*x*>, i. e., $(\langle x \rangle - \sigma) < x_\mathrm{i} < (\langle x \rangle + {} + \sigma)$; 95 % are within two standard deviations, and only 0.3 % are farther than 3σ from <*x*>. So, for example, if an experimental data point lies 3σ from a theoretical prediction, there is a strong chance that either the prediction is not correct or there are systematic errors which affect the experiment. Result of the measurement of *E* can also be reported as, $E = \langle x \rangle \pm \sigma$.

2.2. Experiment 1. The Determination of Gravitational Acceleration

Purpose:

The purpose of this experiment is to measure the earth's gravitational acceleration from an object in free fall. You will use the equation of motion of an object in free fall, starting from rest ($v_0 = 0$):

$$y(t) = y_0 + v_0 t + \frac{1}{2} a t^2. \qquad (2.2.1)$$

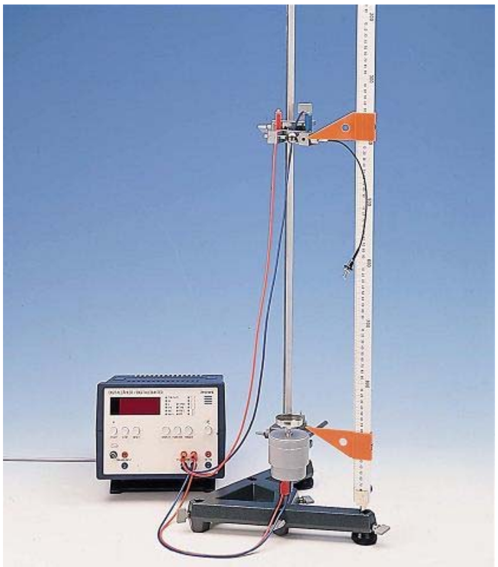

Fig. 27. Equipment: Ball bearing, timer, clamping post, meter stick

Experiment:

Using the apparatus shown above, drop a ball

- 10 times, from;
- 4 different heights (you pick the heights).

To clamp the ball, push in the dowel pin until the ball is tightly clamped, and then tighten the thumbscrew. To release the ball, open the thumbscrew. Be sure to reset the timer to zero before dropping the ball! The lab instructor will show you how to do all this in detail. Be careful when you measure the height: measure from the bottom of the ball to the pressure pad.

Collect your data at four different heights and enter the data in the tables below. For each measured *y* and *t*, you can calculate *g*. Since the initial velocity v_0 is zero, the initial position is h ($y_0 = h$), and the acceleration is $-g$ ($a = -g$), you can rearrange the equation (2.2.1).

To solve for g:

$$g = \frac{2h}{t^2}. \tag{2.2.2}$$

You can do this with your calculator, or put the time values into Excel, and make an equation to solve for g at each value of t and y. For each height, calculate the mean value of gravity and time ($<g>$, $<t>$), and the standard deviation (σ_g, σ_t).

Table 9

Time and Gravity Data

$h_1 =$		$h_2 =$		$h_3 =$		$h_4 =$	
t	g	t	g	t	g	t	g
$<t_1> =$		$<t_2> =$		$<t_3> =$		$<t_4> =$	
$\sigma_t =$		$\sigma_t =$		$\sigma_t =$		$\sigma_t =$	
$<g_1> =$		$<g_2> =$		$<g_3> =$		$<g_4> =$	
$\sigma_g =$		$\sigma_g =$		$\sigma_g =$		$\sigma_g =$	

You now have four mean values of g, each calculated at different height. What is the percent error of each of the values, with respect to the accepted value of $g = 9.81$ [m · s^{-2}]? What is the percent difference between $<g_{\text{highest}}>$ and $<g_{\text{lowest}}>$.

Questions:

1. Discuss the sources of error in the experiment. Were there sources of random error? What were they? What about systematic error?
2. Would you expect the values of g measured at the two different heights to be the same? Why or why not?

3. Suppose you had used a metal ball of a different mass (say, ten times heavier). Would you expect the value of g to be the same, or different? Why?

4. For measuring the distance that the ball drops, we suggested that you measure from the bottom of the ball to the pressure pad, in order to get the most accurate distance. Why shouldn't you measure from the middle of the ball to the pressure pad? Explain.

5. Which of your values for <g> was more precise? Which was more accurate? Explain.

2.3. Experiment 2. Newton's Second Law: the Atwood Machine

Purpose:

To predict the acceleration of an Atwood Machine by applying Newton's 2nd Law and use the predicted acceleration to verify the equations of kinematics with constant acceleration.

The Atwood Machine consists of a pulley of negligible mass and friction over which two masses are suspended (Fig. 28). When the suspended masses are unequal, the system will accelerate in the direction of the larger mass. In this experiment you will measure the acceleration and compare to the acceleration predicted by Newton's 2nd Law. For the purpose of this experiment, we will consider the acceleration to be constant. The system will begin at rest, at position y above the table. You will measure the distance y and the time t required for the system to fall to the table. The system's acceleration can then be calculated using kinematics equations.

Procedure:

Use a length of string such that, when one mass holder is on the table, the other is between 50 cm and 60 cm above the table. Make sure that one mass holder is directly in front of the meter stick.

Measure the initial mass on each holder, including the holder.

Record the initial values:

m_1________________[g]

m_2________________[g].

These numbers should initially be (approximately) equal.

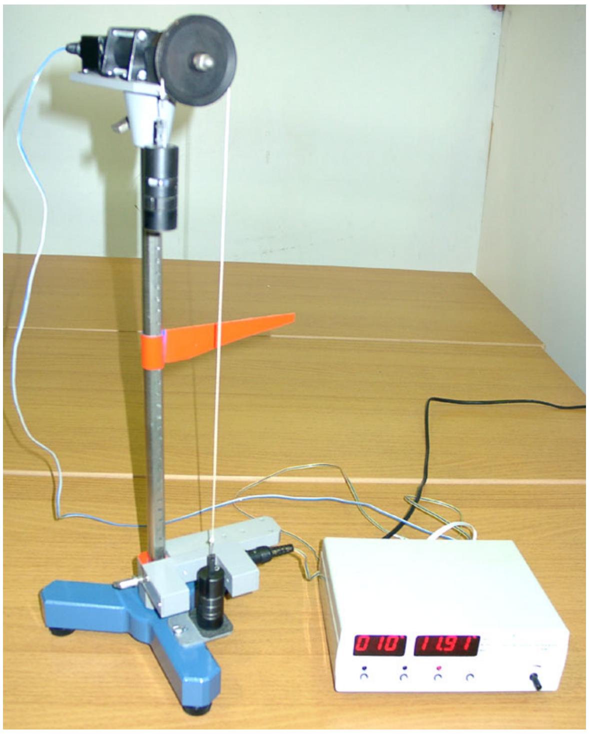

Fig. 28. Equipment: m_2 on the right, m_1 on the left

While gently holding the system (place your finger under the mass holder), obtain a difference of 1 gram between the sides. Let go of the mass holder to see if the system moves. If the system does not move, see if it will move if you very gently tap the larger mass. If the system still does not move, continue adding masses and tapping the heavier mass until the system does move. Record the additional mass required to start the system moving.

1. With an equal total mass on each side, remove a 10 gram mass from the side farthest from the meter stick (m_2) and add it to the side in front of the meter stick, m1, thus making the mass difference between the two 20 [g].
2. Pull m_2 (the light side) down to the table and hold it in place. Read the distance of m1 (the heavy side) above the table by sighting across the bottom of the mass holder to the meter stick.
3. Record this distance in the data table as y.
4. Release the lighter mass; the heavier mass will then fall to the table.
5. Use a stopwatch to determine the time required for the heavier mass to fall.
6. Record the time in the data table as t. Perform a total of five trials.

Return the 10 [g] mass to m_2, remove the 20 [g] mass from m_2, and add it to m_1 so that m1 is 40 [g] heavier than m_2. Repeat steps 2–6.

Add the 10 [g] mass to m_1 from m_2 so that m_1 is 60 [g] heavier than m_2. Repeat steps 2–6. You should now have three sets of data, each having five values for y and t.

Data:

Mass required to start the system moving:________________.

SET 1

Trial	y [m]	t [s]	a_y [m/s]
A			
B			
C			
D			
E			

SET 2

Trial	y [m]	t [s]	a_y [m/s]
A			
B			
C			
D			
E			

Using the equation:

$$y = \frac{1}{2} a_y t^2, \qquad (2.3.1)$$

and the values that were obtained for y and t; compute five values of a_y for each of the data sets. Don't forget to change y from cm to m in your equation. Compute the average value of a_y for each of the data sets. These will be taken as the experimental values of acceleration. Compute the standard deviation of a_y for each of the data sets. Apply Newton's 2nd Law to an Atwood's Machine and derive a formula for the expected acceleration in terms of m_1 and m_2. Start by making a free body diagram in the box below. The instructions following that diagram will help you find the theoretical equations for a_y. Consider each mass as a separate object and draw a free body diagram for each. Note that all forces act in the y-direction.

Write $\sum F_y = ma_y$ for each of the mass obtain two linear equations that include the acceleration of each mass. Solve the resulting system of linear equations to obtain a theoretical value for a_y. Note that the masses are constrained to move together, so $a_{1y} = a_{2y} = a_y$! (Note that absolute value signs may be needed in this equation, depending on what axes you are using. Using your values for m_1 and m_2, compute the expected acceleration for each of your three trials. These will be taken as the theoretical values of acceleration. Compute the % error between the experimental and the accepted values of acceleration. Summarize your results.

Questions:

1. The pulley is not, in fact, frictionless and massless. At the beginning of the lab you found the mass difference needed to start movement of the system. How can this data be used to approximate the effect of friction?
2. What are possible sources of error in measuring the values of t and y? What effect will these errors have on your results? Suggest a possible change to the procedure that could eliminate these errors.
3. Which data set in produced the most accurate value of a_y? Why?
4. What value of a_y would Newton's 2nd Law predict as m_1 becomes much larger than m_2?
5. Why would this value be expected? Hint: consider: $\lim_{m_1 \to \infty} \frac{m_1 - m_2}{m_1 + m_2}$.

2.4. Sound Waves: Harmonic in an Open-Ended Pipe

Purpose:

To determine the speed of sound in air using a water-filled, open ended pipe, a tuning fork, and the knowledge of harmonics.

Theory:

Sound waves are longitudinal waves, meaning that the direction of the displacement or movement of the particles is in the *same direction* as the propagation of the wave. Interference can occur in sound waves: when there is constructive interference of reverberating waves, the result is increased sound intensity. If a tuning fork is struck, the fork vibrates, causing a compression of the air particles moving spherically outward from the fork. When the fork oscillates inward, it creates an area of rarefied air particles. A region of compressed air followed by a region of rarefied air constitutes one wavelength. If the fork is struck over a tube filled with water, a sector of this wave will be captured by the tube. When the air particles hit the surface of water, most of energy is reflected at the water's surface due to the differences in the densities of the air and the water.

With each vibration of the tuning fork, another spherical wave front moves out from the fork, and a part of the new wave is also captured by the tube. When the height of the water is adjusted so that an *incoming* (compressed air) wave front is just meeting a *reflected* (compressed air) wave front at the open end of the tube, the waves will interfere constructively. When this happens over and over again each second, our ears can detect this constructive interference as increased sound intensity: it gets loud!

Using our simple tube and a tuning fork, we can discover the speed that sound travels in air. We know that:

$$v = f\lambda. \qquad (2.4.1)$$

The velocity of any wave is directly proportional both to the wavelength λ of the sound wave created by the tuning fork and f, the frequency of vibration of the tuning fork. The frequency is printed on the tuning fork. We can determine the wavelength by knowing a little about harmonics, and by creating harmonics with our tube and tuning fork.

Harmonics: the surface of the water is considered a *node*. In transverse waves, the node occurs when the displacement is zero. Likewise, at the water's surface at which the wave is reflected, the net displacement of the wave is zero. At the open end of the tube, the air is free to move, and this point is an *antinode*. Multiple nodes and antinodes will exist in the tube. L is the distance from the top of the tube to the surface of the water, which we will measure with a meter stick. The first harmonic is the case in which there is only one node and one antinode in the tube, and this is also called the fundamental *mode*. In the case of an open ended tube, one wavelength of the fundamental *mode* is four times the length of the tube from the top of the tube to the water's surface:

$$\lambda = 4L. \qquad (2.4.2)$$

Substituting into (2.4.1)

The frequencies of the third and fifth harmonics are $3f_1$, $5f_1$... That is, in a pipe closed at one end and open at the other, only odd harmonics are present. These are given by

$$f_n = nv/4L \quad n = 1, 3, 5, \dots \tag{2.4.3}$$

Note: The diagrams in the textbook are helpful, but also somewhat misleading, as the diagrams are drawn as though the waves are transverse. Imagine regions of compressed air molecules followed by rarefied air moving through the tube to get a better visual picture.

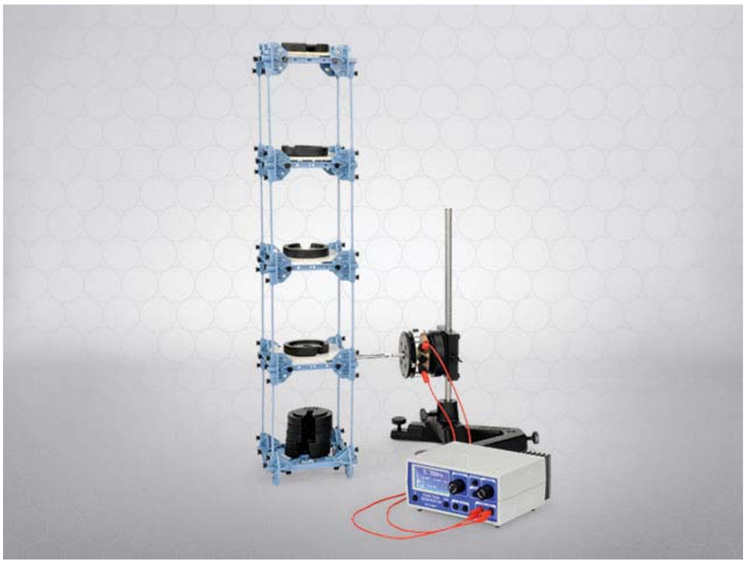

Fig. 29. Equipment: Transparent open-ended tube

Procedure:

1. Secure the tube vertically to the ring stand. Clamp the tuning fork just above the tube. Attach amplifier to the tube, if an amplifier is available.
2. Fill pitcher with water. If your tube does not have a valve, have one partner hold their thumb over the bottom of the drain tube. Pour water into the tube to within 5 cm of the top.
3. Strike tuning fork over the top of the tube.
4. First run: Drain water quickly from the tube, and listen closely for the sound intensity to increase. Make note of the approximate height of the water when that first "harmonic" is heard.
5. Continue to drain the water rather quickly, and a second harmonic will be heard. Make note of the approximate height of the water.
6. Second run: Refill the tube with water.

7. Quickly run the water out to five or six cm above where the first harmonic was heard. Then allow the water to slowly drain, listening for the amplitude of the sound to maximize. Record the height of the water when the sound peaks.

8. You may strike the tuning fork again.

9. Again, run the water down quickly to within five or six cm of the point of the second harmonic. Then drain the water slowly, and record the height of the water when the amplitude of the sound peaks again.

10. If you are careful, you may be able to find a third harmonic.

11. Repeat the procedure using two other tuning forks.

The speed of sound can be determined by making a graph, plotting f versus T, with f on the ordinate (y axis) and T on the abscissa (x axis). For each tuning fork you used, plot that frequency and the corresponding average wavelength calculated in your tables. After plotting these three points on the graph, find the best line to fit the data. The slope of the line is the velocity of sound.

Compare what you discovered with the known velocity of sound in air of 343 $[\mathrm{m} \cdot \mathrm{s}^{-1}]$.

Questions:

1. Did the observed frequency match your prediction, above? If so, how? If not, why not?

2. What is vibrating when these laboratory instruments make sound?

2.5. Electric Fields and Equipotential Surfaces

Purpose:

The purpose of this lab is to explore the electric force per unit charge as a function of the distance from various charged electrode configurations.

Equipment:

Multimeter, apparatus for mapping equipotentials, graph paper (provided). The electric force per unit charge is called the electric field intensity or simply the electric field (E). The electric field is a vector quantity given by

$$\vec{E} = \frac{\vec{F}}{q_0} = k\frac{q}{r^2}r. \qquad (2.5.1)$$

Like all other vector quantities, it has both magnitude and direction. As discussed in the lecture, electric field lines flow from positively to negatively charged regions (positive to ground in this experiment). From the equation above (Coulomb's Law) you should also realize that the magnitude of the electric field decreases as the inverse square of the distance from the point source (in this experiment, the electrodes). This implies that the density of electric field lines (how close together they are) will decrease as you get further away from the source.

Experiment:

Select two configurations. Make sure at least one of them is the parallel plate configuration. Sketch two configurations of the electrodes (the copper shapes on the apparatus) on the graph paper.

Connect the apparatus to the power supply. Turn the power supply on and set it to 18 volts. Connect the ground from the voltmeter to the ground from the power supply. Now, using the positive probe you will mark out some equipotential surfaces around the electrodes. Being careful to not touch the grid with anything other than the probe, locate 10 points each for the following voltages; 3, 6, 9, 12, 15 [V]. Make sure the points are spread out, so that you can get a good sampling of the space around each electrode. Connect the points of equal voltage (potential) with a smooth line and label them. These are equipotential surfaces. Now draw the corresponding electric field lines, with arrows to show the direction of the field.

Questions:

1. In your own words, explain why equipotential surfaces are perpendicular to electric field lines. (Hint: Check your book.)
2. What condition must exist to have a region of nearly uniform electric fields? (Hint: think about your parallel plate configuration.)
3. Where are regions of strongest and weakest electric fields located? (Hint: Re-read the introduction.)
4. Can electric field lines ever cross? Explain. (Hint: Remember that the electric field is a vector.

2.6. Ohm’s Law, Measurement of Voltage, Current, Resistance

Purpose:

In this experiment you will learn to use the multi-meter to measure voltage, current and resistance.

Equipment:

Bread board, variable DC power supply, resistors (15 kΩ, 22 kΩ, 47 kΩ, 68 kΩ, and 1 MΩ), multi- meter, ammeter.

The measurements of voltage, current, and resistance that you will make will be made using direct current (D.C.). D.C. refers to direct current which flows in only one direction down a wire. Usually it is a steady current, meaning that its magnitude is constant in time. "D.C." can also be used to refer to voltage. Of course, unlike current, voltage does not "flow". Instead, "D.C. voltage" (or "D.C. potential") means a constant voltage which has only one polarity. One of the major concepts that will be used in this experiment is Ohm's law, which we discussed in lecture. This law states the relation among the three quantities voltage, current, and resistance:

$$V = IR, \tag{2.6.1}$$

where I is the current measured in units of amperes, [A], V is the voltage in units of volts, [V] and R is the resistance in units of Ohms, [Ω].

An easy way to think of this law is to imagine that a "current" flows through a wire just like water flowing through a pipe; the narrower the pipe, the greater the resistance. Fig. 30 shows the standard symbols for showing a battery and a resistor in a circuit. Remember that current flows from positive to negative, representing the flow of positive charge in the wire. (Remember also that it is really the negatively charged electrons that actually do the moving!) In reality, any circuit element (like a light bulb) can act as a resistor. For experiments and for building circuits, small resistors of known resistance can be added to the circuit.

Experiment:

1. Connect the circuit as shown by the diagram in Fig. 30. Use the variable power supply and a 15 [KΩ] resistor. Adjust the power supply voltage to 5 volts.
2. Use the multi-meter to measure the voltage across the resistor (VR).
3. Use the ammeter to measure the current through the resistor.
4. Record your measurements of voltage and current in Data Table.
5. Repeat steps 2 through 4 for 10 different values of voltage, in steps of 1 [V]. Record all measurements in your table.

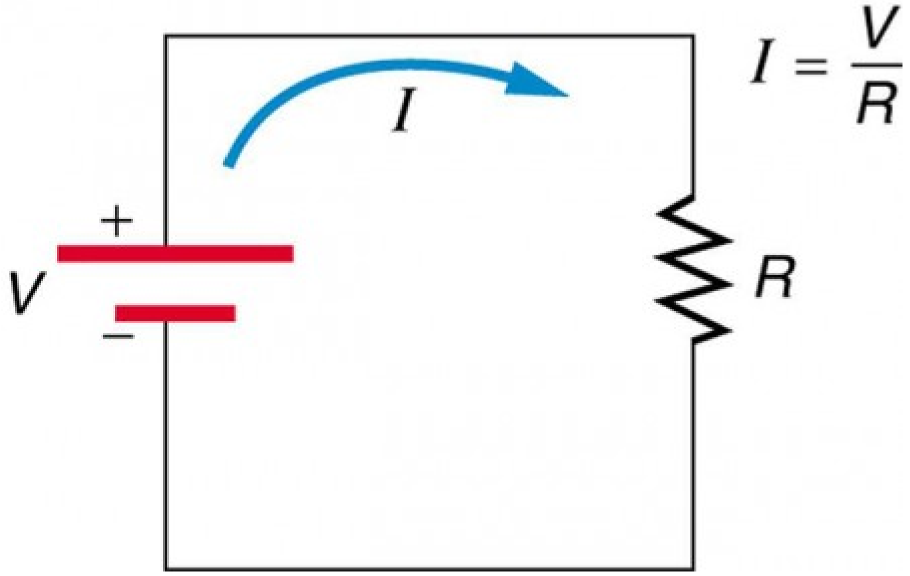

Fig. 30. Diagram showing a voltage forcing a current to flow through a resistor

Questions:

1. In Part 2 of the experiment, was the value of the resistance you obtained from your graph within the tolerance given by the resistor's color code? Explain.
2. List some possible sources of error that might have affected your measurements in Part 2 of the experiment.
3. Is a plot of current vs. voltage ALWAYS a straight line? Explain why or why not.
4. Sometimes, people might plot I on the y-axis and V on the x-axis. (In this case, the slope will be 1/R.) Why might a scientist want to plot the data that way?

2.7. Kirchhoff's Laws

Purpose:

To verify Kirchhoff's Laws by comparing voltages obtained from a real circuit to those predicted by Kirchhoff's Laws.

A simple circuit is one that can be reduced to an equivalent circuit containing a single resistance and a single voltage source. Many circuits are not simple and require the use of Kirchhoff's Laws to determine voltage, current, or resistance values. Kirchhoff's Laws for current and voltage are given by equations 1 and 2:

$$\sum_{Junction} I = 0, \tag{2.7.1}$$

$$\sum_{Loop} \Delta V = 0. \tag{2.7.2}$$

In this experiment, we will construct two circuits with 4 resistors and a voltage source. These circuits will not be simple, thus Kirchhoff's Laws will be required to determine the current in each resistor. We will then use a digital multi-meter to obtain an experimental value for the voltage across each resistor in the circuits. Kirchhoff's Laws will then be applied to the circuits to obtain theoretical values for the current in each resistor. By applying Ohm's Law, we can then obtain a theoretical value for the voltage across each resistor. The experimental and theoretical voltages can then be compared by means of % error.

Equipment:

Proto-board, 4 resistors: ($R1$ = 68 kΩ, $R2$ = 47 kΩ, $R3$ = 15 kΩ, $R4$ = 1000 kΩ), Digital multi-meter Variable power supply, Wire leads and alligator clips.

Procedure:

1. Using the proto-board, the 4 resistors, the variable power supply, and the wire leads and alligator clips; construct the circuit.
2. Turn on the power supply. Connect the multi-meter across the power supply and adjust the voltage to 8.0 volts.
3. Connect the multi-meter across each of the 4 resistors. Record these 4 values of voltage in the data table.
4. Turn the power supply off and disconnect the circuit.
5. Add a second power supply to the circuit.

Turn on the power supplies. Adjust the voltages V0 and V1 to 4.0 volts.

6. Connect the multi-meter across each of the 4 resistors. Record these 4 values of voltage in the data table.

Analysis:

Turn the power supply off and disconnect the circuit. Use equations (2.7.1) and (2.7.2) to write a system of linear equations that may be solved for the current in each branch of the circuit. Then, solve the system to obtain a theoretical value for each current. Using the currents obtained in step 1 of the analysis; apply Ohm's Law to determine the theoretical voltage across each resistor. Compare the theoretical voltages obtained in step 2 of the analysis to those measured in the actual circuit. Repeat steps 1 to 3 for the second circuit. Record the theoretical voltages, the experimental voltages, and the % errors in the results table.

2.8. AC Voltages, Frequency and the Use of the Oscilloscope

Purpose:

After completing this experiment, you will be able to use the oscilloscope to measure unknown voltages and frequencies.

Equipment:

Oscilloscope, signal generator, connecting wires.

Procedure:

In this experiment, we will try a few different things with the oscilloscope to help you become more familiar with it. Turn your oscilloscope on and let it warm up for a few seconds. While it's warming up check the following switches to make sure they're set correctly. Look for the area marked Trigger. There is a switch marked Mode which should be set to Auto. There's also a switch here marked Source. This switch should be all the way up, (it will be set on line). We will also be using Channel 1 throughout this lab, unless otherwise directed, so any switches marked CH1 or CH2 should be set to CH1. For the gray oscilloscope, leave the Source set to CH1. By now your oscilloscope should be warmed up. If you do not see something on the screen, try pushing the Beam Find button and use the Position knobs to move the trace back to the center of the screen. If you still see nothing, try turning the Intensity up a little. You will notice the screen is divided into squares. Each square is called a division. This is important to know because you will notice the 3 largest knobs are marked sec/div or volts/div. So when we see that something has moved say 3 divisions, we will be able to convert that to Volts or seconds. This is how we will make our measurements. Now, look for the knob marked sec/div. This knob controls the «Time base». Turn this knob to .1 sec/div. For the grey oscilloscope you may need different values for this knob. You can now see the trace move across the screen from left to right. You will notice it takes about a second to cross the screen. Actually it takes exactly a second to cross the screen. We can calculate this because the time base knob is on .1 sec/div and the grid on the screen is divided into 10 divisions horizontally and 8 divisions vertically. Multiplying gives:

$$.1\frac{\text{sec}}{\text{div}} \cdot 10 = 1\text{ second.} \tag{2.8.1}$$

If you now set the time base knob to something smaller like 1ms/div, (Remember, m stands for milli = 10-3), you will see the trace move so fast that it's hard or impossible to see. Remember, the horizontal axis of the screen is where we measure time, so you just have to count the divisions between any two points you're interested in and multiply by the scale on the sec/div knob. So now you can (or will soon be able to) find how long it takes the trace to move between any two points on the screen. The vertical axis works the same way, except we use it to measure volts rather than time. A helpful way to think about these knobs is that changing the setting is like changing which units you can use on a ruler; if you're measuring a piece of paper you can use centimeters or millimeters or whatever. But the object you're measuring doesn't change as you switch between scales. Or you might imagine you're zooming in or out with a variable magnifying glass; changing the knobs just changes the magnification, and the numbers tell you what the magnification scale is. Connect a triangular wave of about 1 KHz frequency from the signal generator to CH 1 of the scope. Since we are measuring a wave whose frequency is in the 1 kHz range, the typical period will be in the ms range. Set the time base knob to 1 ms/div. Make sure the cal knob is all the way clockwise. Set the volts/div switch to 2 volts/div. Check that the small knob marked cal on the volts/div knob is turned all the way clockwise, i.e. in its calibrated position. If this knob is not in the calibrated position, your measurements will be incorrect. Adjust the amplitude of the signal generator using the output level knob to give you a +/–6 volt wave (12 volts peak to peak). Be sure that the DC offset knob on the generator has the white dot pointed up, i.e. zero volts DC offset. You should see a triangular wave which has a positive maximum three divisions above the middle line across the screen and a negative minimum three divisions below. The sweep (i. e. the movement of the dot across the screen) is "triggered" when the voltage crosses the threshold set with the level control. The point at which this crossing takes place then becomes the "origin" of the graph. The threshold is controlled with the small knob marked level. Use the horizontal position knob to move the whole trace to the right so you can see the beginning. Explore what happens when the level knob is rotated. Note that you can change the location of the origin of your graph.

Now look at the button or switch right by the level control marked slope. This controls the polarity of the voltage for which the triggering takes place. Both positive triggering and negative triggering are possible.

Explore what happens when this button is both up and down or in and out for the gray oscilloscope. You may have to adjust the level to get stable triggering after this. Now make an accurate measurement of the frequency of the triangular wave. You will measure the period, *T*, the time between any two repeating points on the wave. For example, count accurately the divisions between the two maxima, and then use the information about sec/div from the time base knob setting to find *T*. The frequency is given by:

$$f = \frac{1}{T}. \tag{2.8.2}$$

Question:

1. How many seconds does it take the trace to go across 1/2 the screen at the 1 ms/div setting?
2. Is the signal generator frequency knob very accurate?

2.9. RC Circuits

Purpose:

In this experiment you will investigate the interaction between current and magnetic fields. You will determine the direction of the *B* field surrounding a long straight wire using a compass (Oersted's experiment), find the induced voltage in a small inductor coil, and show that the magnitude of the *B* field decreases as 1/r, determine the permeability of free space using a Hall probe and the constant magnetic field near a long straight wire.

Equipment:

DC power supply, function generator, oscilloscope, inductor coil, small compass, and long straight wire apparatus.

When a current *I* exists in a long straight wire, a magnetic field *B* is generated around the wire. The field lines are concentric circles surrounding the wire. The magnitude of the magnetic field (*B*) as a function of *I* and the distance (*r*) away from the wire is given by:

$$B = \frac{\mu_0 I}{2\pi r}, \tag{2.9.1}$$

where $\mu_0 = 4\pi \cdot 10^{-7}$ [Tm/A], I is in Amperes, r is in meters, and B is in Tesla. (The direction of B, of course, is given by the right hand rule. (Note that this equation is actually derived assuming that the long straight wire is actually infinitely long!!). If the current in the long straight wire is constant in time, the B field created by that current will also be constant in time. In this case, the direction of the B field can be determined by observing its effect on a small compass placed in the vicinity of the long straight wire. This is basically Oersted's experiment. If the current in the long straight wire is an alternating current produced by a sine wave generator, the B field surrounding the wire will also be time-varying. A changing magnetic field can induce a current in a wire, because it induces an electromotive force. This is Faraday's law, and is part of the endless hall of mirrors of reciprocal interactions between electricity and magnetism that we have been emphasizing in class. Faraday's law states that the induced emf in a coil of wire (in this case, that's the "inductor coil") placed near the long straight wire is

$$\Theta = \frac{\Delta\Phi}{\Delta t}, \tag{2.9.2}$$

where ΔΦ is the magnetic flux, which can be changed by changing the magnetic field. (The flux can be changed by a few other things too, which we will discuss in class!) So, if the magnetic field going through the inductor coil is changing, alternating in magnitude and direction because of the sine-wave generator, an alternating voltage will be induced in the wire. In other words: The current in the long wire oscillates because it is coming from a sine wave generator….which makes the B field around the wire oscillate….which makes the induced emf in the small "inductor coil" oscillate too! (Which makes an oscillating current in the inductor coil… And yes, the current in the inductor coil will generate a tiny little B field of its own…). According to Faraday's law, this induced voltage in the coil is proportional to the rate of change of the magnetic flux through the coil, and hence to the magnitude of the time-varying B field. Therefore, a measurement of the voltage induced in the coil, as the coil is placed at different distances from the wire, provides a relative measure of the magnitude of the B field at different distances from the wire. Note that the quantity actually measured is an alternating electric voltage, but its magnitude is proportional to the *B* field and will be taken to be a relative measurement of the *B* field at a given point. In other words, we are not measuring *B* directly. We are measuring the emf caused by *B*, and by measuring the emf at different distances r, we can infer how *B* changes as a function of distance.

Experiment:

1. Connect the circuit, using the direct current power supply. Stand the long wire apparatus on its end so that the long wire is vertical.
2. Turn on the power supply. (Notes: the DC power supply you will use for this experiment is the power supply needs to be set to the maximum voltage).
3. Place the compass on the platform at various positions around the wire, and record the direction of the compass needle at each position. Record your measurements in Data Table.

Questions:

1. Explain how the earth's magnetic field could affect your results. Based only on your data in Data Table above, can you tell what side of the laboratory is facing (magnetic) North?
2. Use the direction of the compass needles and the right hand rule to determine whether the current in the wire is going up or down.
3. Why does the plot of "*B*" vs. 1/*r* look like a straight line?
4. When the direct current is 2.00 [A] in a single wire of the bundle of 10 wires, the total current in the bundle of wire that approximates the long straight wire is 20.0 [A]. What is the magnitude of the *B* field 3.00 cm from this long straight wire carrying a current of 20.0 [A]? What is the magnitude of the *B* field 9.00 [cm] from the wire carrying 20.0 [A]?
5. A constant current is in a long straight wire in the plane of the paper in the direction shown below by the arrow. Point *X* is in the plane of the paper above the wire, and point *Y* is in the plane of the paper but below the wire. What is the direction of the *B* field at point *X*? What is the direction of the *B* field at point *Y*?

2.10. Reflection and Refraction

Purpose:

The purpose of this experiment is to investigate two of the basic laws of optics, namely the law of reflection and Snell's law.

Reflection and refraction are two commonly observed optical properties of light. Whenever a light strikes the surface of some material at an angle, part of the wave is reflected and part is transmitted (or absorbed). Due to refraction, the velocity of transmitted light is less than the velocity before it entered the medium. The denser the medium, the more the light is slowed down. This is due to interaction between the light and the orbiting electrons in the atoms comprising the material. When light travels from one material into another, it not only may change velocity, but it may be bent at a different angle in the new medium than the angle at which it entered. Snell's law states the relationship between the indices of refraction of the two materials, and the light's angle of incidence and angle of refraction:

$$n_1 \sin \alpha = n_2 \sin \beta, \tag{2.10.1}$$

Equipment:

ray box, semicircular solid block, semicircular hollow plastic block, polar graph paper, tape.

Experiment:

1. Block off all slits on the ray box with masking tape, except for the center slit.
2. Place the solid semicircular block so that it is centered on the polar graph paper with its flat edge facing the ray box. Secure the block to the graph paper with tape so that the flat side lies along the center line (parallel to the short side of the paper). The center of the block should be aligned with the center of the coordinate system.
3. Angles will be measured with respect to the zero degree line on the polar graph paper. Note that all angle measurements will be in the range $0° \leq \theta \leq 90°$.
4. Starting with the light ray at normal incidence (perpendicular) to the flat edge of the block, rotate the graph paper from 0° to 90° in 10° increments, each time recording the angle of reflection and the angle of refraction as the light exits the block through its curved side. Be sure that the light enters the block at the center point. Record your measurements in Table. (Note: the amount of reflected light may be very small, so the reflected trace may be very faint. You will need to dim the room lights to record the angle of reflection.)

Starting with the light ray at normal incidence (perpendicular) to the curved edge of the block, rotate the graph paper from 0° to 90°, each time noting the angle of reflection and the angle of refraction as the light exits the block through its flat side. Be sure that the light exits the block at the center point. Find the angle at which the refracted ray totally disappears. This is the angle at which total internal reflection occurs. Repeat, but using the hollow semicircular block filled with water. Enter your measurements in Table.

Questions:

1. Using Snell's Law, compute the value of the index of refraction of the block. For the refractive index of air, use the value of 1.0. Compute n for the block for each 10° increment from 10 to 80 degrees, and compute the average value. Enter your calculated values in the fourth column of each table.
2. Compare your calculated value of the average index of refraction from Table. Are the values different? Would you expect them to be?
3. Why do you need to make sure that the light enters the block normal to the curved side, and exits the flat side at the center point?
4. In the second orientation of the block (curved edge facing the light box) at what observed angle does the refracted light ray disappear? Is this angle the same for the experiment air as for water? This is the critical angle at which total internal reflection occurs, and is given by the equation $\theta_{\text{critical}} = \sin^{-1}(n_2/n_1)$, where n_1 is the index of refraction of the medium the light is leaving (in our experiment, either plastic or water), and n_2 is the index of refraction of the medium the light is entering (in our case, this is air, and you can assume that $n_2 = 1$). Compute θ_{critical} using your average values of n in Table. How does your calculated value compare with your measured value?

2.11. Diffraction, Wavelength, and Atomic Line Spectra

Purpose:

In this experiment, we will look at the diffraction of light, and how wavelengths can be calculated from diffraction. We will also look at atomic line spectra, which actually relate to the quantum-mechanical energy levels that electrons occupy around the nucleus of an atom; the wavelengths of emitted light from an atom relates to these energy levels.

The wavelengths of the emission lines in the spectrum are given by the Rydberg formula:

$$\frac{1}{\lambda} = RZ^2\left(\frac{1}{n_{\text{final}}^2} - \frac{1}{n_{\text{initial}}^2}\right), \tag{2.11.1}$$

where $R = 1.097 \cdot 10^7$ m^{-1}, Z is the atomic number of the atom, n_{final} is the principal quantum number of the final (lowest) energy state of the electron, and n_{initial} is the principal quantum number of the initial (highest) energy level of the electron. For various values of n_{final}, different "series" of spectral lines occur. Specifically, for hydrogen, n_{final} =1, 2, 3, 4 gives the Lyman, Balmer, Paschen and Brackett series, respectively. The energy of a photon is related to the corresponding wavelength of light as follows $E = hf = h(c/\lambda)$.

The shortest wavelengths of light emitted correspond to the greatest loss of energy and occur when the electron falls the greatest possible number of energy levels. This will happen when the initial energy level is the highest one possible, or $n_{\text{initial}} = \infty$. The longest wavelengths of light emitted corresponds to the smallest loss of energy and occur when the electron falls the least possible number of energy levels. This will happen when the initial energy level is just one principal quantum number above the final level.

Experiment:

1. The TA will show you various gas tubes. Each tube contains a different element that, after being excited by running high voltage through it, will emit different spectral lines, which you can view through the diffraction gratings. This happens because the high voltage makes the electrons in the gas jump up to higher energy levels; light is emitted when the electrons "fall" back down again, emitting photons as they fall. Your task is to draw the lines you see in the demo in the boxes in Table, at the locations where they are observed. Once you have been shown all the gas tubes and sketched the spectra you observe, compare your drawing to the diagrams provided in order to identify which element was in each tube element in each tube. Remember that 400 nm is UV light, and 700 nm is infrared.

2. In this part of the experiment, you will look at the interference of light waves, and use the properties of constructive interference to calculate the wavelengths of light in the spectrum you observe. You will

be looking at the spectrum from a regular 150 Watt light bulb. This is a continuous spectrum, so you won’t see isolated lines. Like all waves, light experiences both constructive and destructive interference. Here, we will actually investigate equations for the two different kinds of interference. Constructive interference will occur when the difference in the distance traveled by two light waves is an integer multiple of the wavelength:

$$\Delta path = m\lambda, \text{where } m = 1, 2, 3, \ldots \qquad (2.11.2)$$

Destructive interference will occur when the difference in the distance traveled by two light waves is a half-integer multiple of the wavelength. In other words:

$$\Delta path = \left(m + \frac{1}{2}\right)\lambda, \text{where } m = 1, 2, 3, \ldots \qquad (2.11.3)$$

Procedure:

Your goal is to measure the wavelength of each color in the visible spectrum. You will have to be able to use your trigonometry skills to calculate the angle θ. Feel free to mark the masking tape on the table where each color is located so you can measure all the colors at once. Since measuring the angle correctly is VERY important, make sure to set up the sides of the triangle carefully. The length of the adjacent side is the distance between the light source (the filament in the bulb) and the diffraction grating. The length of the opposite side is the distance from the filament to where you see the band of light with the color you are measuring. Place the light bulb exactly one meter from the edge of the table. It is important to be lined up with the light bulb to measure exactly one meter. (This means that the length of the adjacent side of the triangle is just equal to 1 meter.) Then place the tape across the table in front of the light bulb. Mark a line on the tape where the filament is located. Turn on the light bulb and look through the diffraction grating, holding the grating at the edge of the table. Again, be sure to line up with the light bulb. Off to the side of the light bulb, you should see a continuous spectrum. Have your lab partner draw a line on the masking tape where you see the middle of each color. Make sure to label each line (green, blue, etc.). Record the distance of each color from the filament (this is the length of the opposite side). Enter your measurements in Table; then calculate the angle, and the wavelength. To calculate the wavelength, you will need to know the distance d between the gratings in the diffraction grating. There are 1000 lines per millimeter, which means that $d = 1000$ nm. Use d in units of nm, so that you will get the wavelength in nm also.

Analysis:

1. Calculate the wavelengths of the colors in the visible spectrum based on your measurements in Part 2 of the experiment; enter the values in the table in Part 2.
2. Calculate the longest and shortest wavelengths for the hydrogen Lyman transitions.
3. Calculate the longest and shortest wavelengths for the hydrogen Balmer transitions.
4. Calculate the longest and shortest wavelengths for the hydrogen Paschen transitions.
5. Based on your calculations for questions, which series (Lyman, Balmer or Paschen) are you observing when you looked at the visible part of the spectrum for atomic hydrogen in Part 1 of the experiment? EXPLAIN!

SUMMARY

This textbook is intended for physics two-year introductory courses requiring algebra and some trigonometry and calculus. This physics textbook is written for students. It is based on the teaching and research experience of numerous physicists and influenced by a strong recollection of their own struggles as students. After reading this book, we hope you see that physics is visible everywhere. Applications present different examples from practical physics.

There is considerable latitude on the part of the instructor regarding the use, organization, level, and content of this book. By choosing the types of problems assigned, the instructor can determine the level of sophistication required of the student. Course Physics is organized such that topics are introduced conceptually with a steady progression to precise definitions and analytical applications. The chapters on modern physics are more complete than many other editions textbooks with an entire chapter devoted to laboratory experiments in physics.

REFERENCIES

1. Jim Breithaupt. Understanding Physics for Advanced Level. Stanley Thornes (Publishers) Ltd. 1990. P. 568.
2. Davis Halliday, Robert Resnick. Fundamentals of Physics. JohnWiley & Sons, Inc. 1970. P. 837.
3. Huge D. Young, Roger A. Freedman, A. Lewis Ford. University Physics with Modern Physics. Pearson. 2012. P. 1516.
4. Irodov I. E. Problems in General Physics. Mir Publishers. 1983. P. 386.
5. Physics. Open Stax College. 2013. URL: http://openstaxcollege.org (дата обращения: 25.12.2020).