

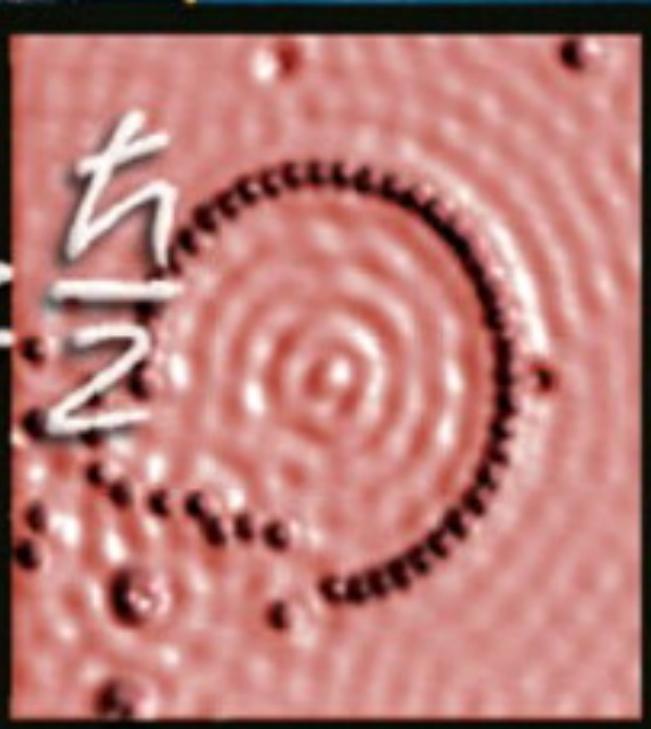
MODERN PHYSICS

FOR SCIENTISTS AND ENGINEERS

SECOND EDITION

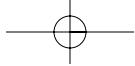


$$E^2 = (pc)^2 + (mc^2)^2$$



$$\Delta x \Delta p \geq$$

John R. Taylor • Chris D. Zafiratos • Michael A. Dubson



RELATIVITY

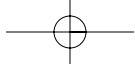
PART



Chapter 1 The Space and Time of Relativity
Chapter 2 Relativistic Mechanics

Two great theories underlie almost all of modern physics, both of them discovered during the first 25 years of the twentieth century. The first of these, relativity, was pioneered mainly by one person, Albert Einstein, and is the subject of Part I of this book (Chapters 1 and 2). The second, quantum theory, was the work of many physicists, including Bohr, Einstein, Heisenberg, Schrödinger, and others; it is the subject of Part II. In Parts III and IV we describe the applications of these great theories to several areas of modern physics.

Part I contains just two chapters. In Chapter 1 we describe how several of the ideas of relativity were already present in the classical physics of Newton and others. Then we describe how Einstein's careful analysis of the relationship between different reference frames, taking account of the observed invariance of the speed of light, changed our whole concept of space and time. In Chapter 2 we describe how the new ideas about space and time required a radical revision of Newtonian mechanics and a redefinition of the basic ideas — mass, momentum, energy, and force — on which mechanics is built. At the end of Chapter 2, we briefly describe general relativity, which is the generalization of relativity to include gravity and accelerated reference frames.



Chapter 1

The Space and Time of Relativity

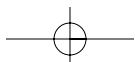
- I.1** Relativity
 - I.2** The Relativity of Orientation and Origin
 - I.3** Moving Reference Frames
 - I.4** Classical Relativity and the Speed of Light
 - I.5** The Michelson–Morley Experiment*
 - I.6** The Postulates of Relativity
 - I.7** Measurement of Time
 - I.8** The Relativity of Time; Time Dilation
 - I.9** Evidence for Time Dilation
 - I.10** Length Contraction
 - I.11** The Lorentz Transformation
 - I.12** Applications of the Lorentz Transformation
 - I.13** The Velocity-Addition Formula
 - I.14** The Doppler Effect*
- Problems for Chapter 1

* Sections marked with a star can be omitted without significant loss of continuity.

1.1 Relativity

Most physical measurements are made *relative* to a chosen reference system. If we measure the time of an event as $t = 5$ seconds, this must mean that t is 5 seconds relative to a chosen origin of time, $t = 0$. If we state that the position of a projectile is given by a vector $\mathbf{r} = (x, y, z)$, we must mean that the position vector has components x, y, z relative to a system of coordinates with a definite orientation and a definite origin, $\mathbf{r} = 0$. If we wish to know the kinetic energy K of a car speeding along a road, it makes a big difference whether we measure K relative to a reference frame fixed on the road or to one fixed on the car. (In the latter case $K = 0$, of course.) A little reflection should convince you that almost every measurement requires the specification of a reference system relative to which the measurement is to be made. We refer to this fact as the **relativity of measurements**.

The theory of relativity is the study of the consequences of this relativity of measurements. It is perhaps surprising that this could be an important subject of study. Nevertheless, Einstein showed, starting with his first paper on relativity in 1905, that a careful analysis of how measurements depend on coordinate systems revolutionizes our whole understanding of space and time, and requires a radical revision of classical, Newtonian mechanics.



Section 1.2 • The Relativity of Orientation and Origin 3

In this chapter we discuss briefly some features of relativity as it applies in the classical theories of Newtonian mechanics and electromagnetism, and then we describe the Michelson–Morley experiment, which (with the support of numerous other, less direct experiments) shows that something is wrong with the classical ideas of space and time. We then state the two postulates of Einstein's relativity and show how they lead to a new picture of space and time in which both lengths and time intervals have different values when measured in any two reference frames that are moving relative to one another. In Chapter 2 we show how the revised notions of space and time require a revision of classical mechanics. We will find that the resulting relativistic mechanics is usually indistinguishable from Newtonian mechanics when applied to bodies moving with normal terrestrial speeds, but is entirely different when applied to bodies with speeds that are a substantial fraction of the speed of light, c . In particular, we will find that no body can be accelerated to a speed greater than c , and that mass is a form of energy, in accordance with the famous relation $E = mc^2$.

Einstein's theory of relativity is really two theories. The first, called the special theory of relativity, is “special” in that its primary focus is restricted to unaccelerated frames of reference and excludes gravity. This is the theory that we will be studying in Chapters 1 and 2 and applying to our later discussions of radiation, nuclear, and particle physics.

The second of Einstein's theories is the general theory of relativity, which is “general” in that it includes accelerated frames of reference and gravity. Einstein found that the study of accelerated reference frames led naturally to a theory of gravitation, and general relativity turns out to be the relativistic theory of gravity. In practice, general relativity is needed only in areas where its predictions differ significantly from those of Newtonian gravitational theory. These include the study of the intense gravity near black holes, of the large-scale universe, and of the effect the earth's gravity has on extremely accurate time measurements (one part in 10^{12} or so). General relativity is an important part of modern physics; nevertheless, it is an advanced topic and, unlike special relativity, is not required for the other topics we treat in this book. Therefore, we have given only a brief description of general relativity in an optional section at the end of Chapter 2.

1.2 The Relativity of Orientation and Origin

In your studies of classical physics, you probably did not pay much attention to the relativity of measurements. Nevertheless, the ideas were present, and, whether or not you were aware of it, you probably exploited some aspects of relativity theory in solving certain problems. Let us illustrate this claim with two examples.

In problems involving blocks sliding on inclined planes, it is well known that one can choose coordinates in various ways. One could, for example, use a coordinate system S with origin O at the bottom of the slope and with axes Ox horizontal, Oy vertical, and Oz across the slope, as shown in Fig. 1.1(a). Another possibility would be a reference frame S' with origin O' at the top of the slope and axes $O'x'$ parallel to the slope, $O'y'$ perpendicular to the slope, and $O'z'$ across it, as in Fig. 1.1(b). The solution of any problem relative to the frame S may look quite different from the solution relative to S' , and it often happens that one choice of axes is much more convenient than the other. (For some examples, see Problems 1.1 to 1.3.) On the other hand, the basic laws of

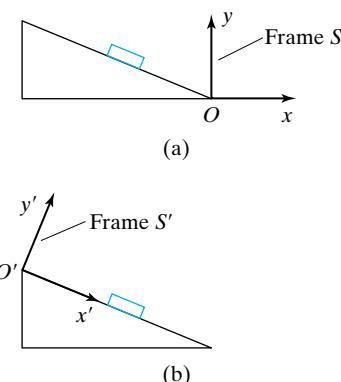
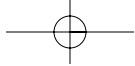


FIGURE 1.1

- (a) In studying a block on an incline, one could choose axes Ox horizontal and Oy vertical and put O at the bottom of the slope.
- (b) Another possibility, which is often more convenient, is to use an axis $O'x'$ parallel to the slope with $O'y'$ perpendicular to the slope, and to put O' at the top of the slope. (The axes Oz and $O'z'$ point out of the page and are not shown.)



4 Chapter I • The Space and Time of Relativity

motion, Newton's laws, make no reference to the choice of origin and orientation of axes and are equally true in either coordinate system. In the language of relativity theory, we can say that Newton's laws are **invariant**, or unchanged, as we shift our attention from frame S to S' , or vice versa. It is because the laws of motion are the same in either coordinate system that we are free to use whichever system is more convenient.

The invariance of the basic laws when we change the origin or orientation of axes is true in all of classical physics — Newtonian mechanics, electromagnetism, and thermodynamics. It is also true in Einstein's theory of relativity. It means that in any problem in physics, one is free to choose the origin of coordinates and the orientation of axes in whatever way is most expedient. This freedom is very useful, and we often exploit it. However, it is not especially *interesting* in our study of relativity, and we will not have much occasion to discuss it further.

1.3 Moving Reference Frames

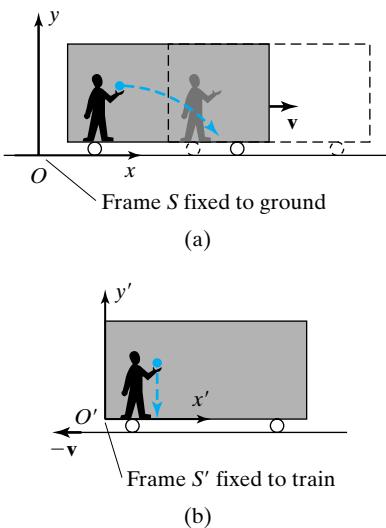


FIGURE 1.2

(a) As seen from the ground, the train and student move to the right; the ball falls in a parabola and lands at the student's feet. (b) As seen from the train, the ball falls straight down, again landing at the student's feet.

As a more important example of relativity, we consider next a question involving two reference frames that are moving relative to one another. Our discussion will raise some interesting questions about classical physics, questions that were satisfactorily answered only when Einstein showed that the classical ideas about the relation between moving reference frames needed revision.

Let us imagine a student standing still in a train that is moving with constant velocity v along a horizontal track. If the student drops a ball, where will the ball hit the floor of the train? One way to answer this question is to use a reference frame S fixed on the track, as shown in Fig. 1.2(a). In this coordinate system the train and student move with constant velocity v to the right. At the moment of release, the ball is traveling with velocity v and it moves, under the influence of gravity, in the parabola shown. It therefore lands to the right of its starting point (as measured in the ground-based frame S). However, while the ball is falling, the train is moving, and a straightforward calculation shows that the train moves exactly as far to the right as does the ball. Thus the ball hits the floor at the student's feet, vertically below his hand.

Simple as this solution is, one can reach the same conclusion even more simply by using a reference frame S' fixed to the train, as in Fig. 1.2(b). In this coordinate system the train and student are at rest (while the track moves to the left with constant velocity $-v$). At the moment of release the ball is at rest (as measured in the train-based frame S'). It therefore falls straight down and naturally hits the floor vertically below the point of release.

The justification of this second, simpler argument is actually quite subtle. We have taken for granted that an observer on the train (using the coordinates x', y', z') is entitled to use Newton's laws of motion and hence to predict that a ball which is dropped from rest will fall straight down. But is this correct? The question we must answer is this: If we accept as an experimental fact that Newton's laws of motion hold for an observer on the ground (using coordinates x, y, z), does it follow that Newton's laws also hold for an observer in the train (using x', y', z')? Equivalently, are Newton's laws invariant as we pass from the ground-based frame S to the train-based frame S' ? Within the framework of classical physics, the answer to this question is "yes," as we now show.

Since Newton's laws refer to velocities and accelerations, let us first consider the velocity of the ball. We let \mathbf{u} denote the ball's velocity relative to the ground-based frame S , and \mathbf{u}' the ball's velocity relative to the train-based S' .

Since the train moves with constant velocity \mathbf{v} relative to the ground, we naturally expect that

$$\mathbf{u} = \mathbf{u}' + \mathbf{v} \quad (1.1)$$

We refer to this equation as the **classical velocity-addition formula**. It reflects our common-sense ideas about space and time, and asserts that velocities obey ordinary vector addition. Although it is one of the central assumptions of classical physics, equation (1.1) is one of the first victims of Einstein's relativity. In Einstein's relativity the velocities \mathbf{u} and \mathbf{u}' do *not* satisfy (1.1), which is only an approximation (although a very good approximation) that is valid when all speeds are much less than the speed of light, c . Nevertheless, we are for the moment discussing classical physics, and we therefore assume for now that the classical velocity-addition formula is correct.

Now let us examine Newton's three laws, starting with the first (the law of inertia): A body on which no external forces act moves with constant velocity. Let us assume that this law holds in the ground-based frame S . This means that if our ball is isolated from all outside forces, its velocity \mathbf{u} is constant. Since $\mathbf{u}' = \mathbf{u} - \mathbf{v}$ and the train's velocity \mathbf{v} is constant, it follows at once that \mathbf{u}' is also constant, and Newton's first law also holds in the train-based frame S' . We will find that this result is also valid in Einstein's relativity; that is, in both classical physics and Einstein's relativity, Newton's first law is invariant as we pass between two frames whose relative velocity is constant.

Newton's second law is a little more complicated. If we assume that it holds in the ground-based frame S , it tells us that

$$\mathbf{F} = m\mathbf{a}$$

where \mathbf{F} is the sum of the forces on the ball, m its mass, and \mathbf{a} its acceleration, all measured in the frame S . We now use this assumption to show that $\mathbf{F}' = m'\mathbf{a}'$, where \mathbf{F}' , m' , \mathbf{a}' are the corresponding quantities measured relative to the train-based frame S' . We will do this by arguing that each of \mathbf{F}' , m' , \mathbf{a}' is in fact equal to the corresponding quantity \mathbf{F} , m , and \mathbf{a} .

The proof that $\mathbf{F} = \mathbf{F}'$ depends to some extent on how one has chosen to define force. Perhaps the simplest procedure is to define forces by their effect on a standard calibrated spring balance. Since observers in the two frames S and S' will certainly agree on the reading of the balance, it follows that any force will have the same value as measured in S and S' ; that is, $\mathbf{F} = \mathbf{F}'$.*

Within the domain of classical physics, it is an experimental fact that any technique for measuring mass (for example, an inertial balance) will produce the same result in either reference frame; that is, $m = m'$.

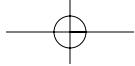
Finally, we must look at the acceleration. The acceleration measured in S is

$$\mathbf{a} = \frac{d\mathbf{u}}{dt}$$

where t is the time as measured by ground-based observers. Similarly, the acceleration measured in S' is

$$\mathbf{a}' = \frac{d\mathbf{u}'}{dt'} \quad (1.2)$$

*Of course, the same result holds whatever our definition of force, but with some definitions the proof is a little more roundabout. For example, many texts define force by the equation $\mathbf{F} = m\mathbf{a}$. Superficially, at least, this means that Newton's second law is true by *definition* in both frames. Since $m = m'$ and $\mathbf{a}' = \mathbf{a}$ (as we will show shortly), it follows that $\mathbf{F} = \mathbf{F}'$.



6 Chapter I • The Space and Time of Relativity

where t' is the time measured by observers on the train. Now, it is a central assumption of classical physics that time is a single universal quantity, the same for all observers; that is, the times t and t' are the same, or $t = t'$. Therefore, we can replace (1.2) by

$$\mathbf{a}' = \frac{d\mathbf{u}'}{dt}$$

Since

$$\mathbf{u}' = \mathbf{u} - \mathbf{v}$$

we can simply differentiate with respect to t and find that

$$\mathbf{a}' = \mathbf{a} - \frac{d\mathbf{v}}{dt} \quad (1.3)$$

or, since \mathbf{v} is constant, $\mathbf{a}' = \mathbf{a}$.

We have now argued that $\mathbf{F}' = \mathbf{F}$, $m' = m$, and $\mathbf{a}' = \mathbf{a}$. Substituting into the equation $\mathbf{F} = m\mathbf{a}$, we immediately find that

$$\mathbf{F}' = m'\mathbf{a}'$$

That is, Newton's second law is also true for observers using the train-based coordinate frame S' .

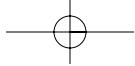
The third law,

$$(\text{action force}) = -(\text{reaction force})$$

is easily treated. Since any given force has the same value as measured in S or S' , the truth of Newton's third law in S immediately implies its truth in S' .

We have now established that if Newton's laws are valid in one reference frame, they are also valid in any second frame that moves with constant velocity relative to the first. This shows why we could use the normal rules of projectile motion in a coordinate system fixed to the moving train. More generally, in the context of our newfound interest in relativity, it establishes an important property of Newton's laws: If space and time have the usual properties assumed in classical physics, Newton's laws are invariant as we transfer our attention from one coordinate frame to a second one moving with constant velocity relative to the first.

Newton's laws would *not* still hold in a coordinate system that was *accelerating*. Physically, this is easy to understand. If our train were accelerating forward, just to keep the ball at rest (relative to the train) would require a force; that is, the law of inertia would not hold in the accelerating train. To see the same thing mathematically, note that if $\mathbf{u}' = \mathbf{u} - \mathbf{v}$ and \mathbf{v} is *changing*, \mathbf{u}' is not constant even if \mathbf{u} is. Further, the acceleration \mathbf{a}' as given by (1.3) is not equal to \mathbf{a} , since $d\mathbf{v}/dt$ is not zero; so our proof of the second law for the train's frame S' also breaks down. In classical physics the unaccelerated frames in which Newton's laws hold (including the law of inertia) are often called **inertial frames**. In fact, one convenient definition (good in both classical and



Section 1.4 • Classical Relativity and the Speed of Light 7

relativistic mechanics) of an inertial frame is just that it is a frame where the law of inertia holds. The result we have just proved can be rephrased to say that an accelerated frame is **noninertial**.

1.4 Classical Relativity and the Speed of Light

Although Newton's laws are invariant as we change from one unaccelerated frame to another (if we accept the classical view of space and time), the same is not true of the laws of electromagnetism. We can show this by separately examining each law — Gauss's law, Faraday's law, and so on — but the required calculations are complicated. A simpler procedure is to recall that the laws of electromagnetism demand that in a vacuum, light signals and all other electromagnetic waves travel in any direction with speed*

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3.00 \times 10^8 \text{ m/s}$$

where ϵ_0 and μ_0 are the permittivity and permeability of the vacuum. Thus if the electromagnetic laws hold in a frame S , light must travel with the same speed c in all directions, as seen in S .

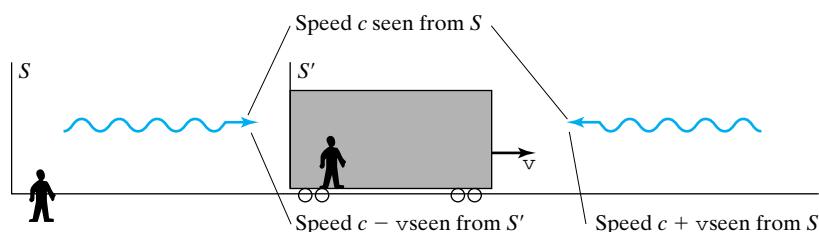
Let us now consider a second frame S' traveling relative to S and imagine a pulse of light moving in the same direction as S' , as shown on the left of Fig. 1.3. The pulse has speed c relative to S . Therefore, by the classical velocity-addition formula (1.1), it should have speed $c - v$ as seen from S' . Similarly, a pulse traveling in the opposite direction would have speed $c + v$ as seen from S' , and a pulse traveling in any other oblique direction would have a different speed, intermediate between $c - v$ and $c + v$. We see that in the frame S' the speed of light should vary between $c - v$ and $c + v$ according to its direction of propagation. Since the laws of electromagnetism demand that the speed of light be exactly c , we conclude that these laws — unlike those of mechanics — could not be valid in the frame S' .

The situation just described was well understood by physicists toward the end of the nineteenth century. In particular, it was accepted as obvious that there could be only one frame, called the **ether frame**, in which light traveled at the same speed c in all directions. The name "ether frame" derived from the belief that light waves must propagate through a medium, in much the same way that sound waves were known to propagate in the air. Since light propagates through a vacuum, physicists recognized that this medium, which no one

Isaac Newton
(1642–1727, English)



Newton was possibly the greatest scientific genius of all time. In addition to his laws of motion and his theory of gravity, his contributions included the invention of calculus and important discoveries in optics. Although he believed in "absolute space" (what we would call the ether frame), Newton was well aware that his laws of motion hold in all unaccelerated frames of reference. From a modern perspective, it is surprising that Newton devoted much of his time to finding ways to manufacture gold by alchemy and to dating the creation of the world (3500 B.C. was his answer) using biblical chronology.

**FIGURE 1.3**

Frame S' travels with velocity v relative to S . If light travels with the same speed c in all directions relative to S , then (according to the classical velocity-addition formula) it should have different speeds as seen from S' .

*More precisely, $c = 299,792,458 \text{ m/s}$. In fact, the determination of c has become so accurate that since 1984, the meter has been defined in terms of c , as the fraction $1/299,792,458$ of the distance traveled by light in 1 second. This means that, by definition, c is $299,792,458 \text{ m/s}$ exactly.

8 Chapter I • The Space and Time of Relativity

had ever seen or felt, must have unusual properties. Borrowing the ancient name for the substance of the heavens, they called it the “ether.” The unique reference frame in which light traveled at speed c was assumed to be the frame in which the ether was at rest. As we will see, Einstein’s relativity implies that neither the ether nor the ether frame actually exists.

Our picture of classical relativity can be quickly summarized. In classical physics we take for granted certain ideas about space and time, all based on our everyday experiences. For example, we assume that relative velocities add like vectors, in accordance with the classical velocity-addition formula; also, that time is a universal quantity, concerning which all observers agree. Accepting these ideas we have seen that Newton’s laws should be valid in a whole family of reference frames, any one of which moves uniformly relative to any other. On the other hand, we have seen that there could be no more than one reference frame, called the ether frame, relative to which the electromagnetic laws hold and in which light travels through the vacuum with speed c in all directions.

It should perhaps be emphasized that although this view of nature turned out to be wrong, it was nevertheless perfectly logical and internally consistent. One might argue on philosophical or aesthetic grounds (as Einstein did) that the difference between classical mechanics and classical electromagnetism is surprising and even unpleasing, but theoretical arguments alone could not decide whether the classical view is correct. This question could be decided only by experiment. In particular, since classical physics implied that there was a unique ether frame where light travels at speed c in all directions, there had to be some experiment that showed whether this was so. This was exactly the experiment that Albert Michelson, later assisted by Edward Morley, performed between the years 1880 and 1887, as we now describe.

If one assumed the existence of a unique ether frame, it seemed clear that as the earth orbits around the sun, it must be moving relative to the ether frame. In principle, this motion relative to the ether frame should be easy to detect. One would simply have to measure the speed (relative to the earth) of light traveling in various directions. If one found different speeds in different directions, one would conclude that the earth is moving relative to the ether frame, and a simple calculation would give the speed of this motion. If, instead, one found the speed of light to be exactly the same in all directions, one would have to conclude that at the time of the measurements the earth happened to be at rest relative to the ether frame. In this case one should probably repeat the experiment a few months later, by which time the earth would be at a different point on its orbit and its velocity relative to the ether frame should surely be nonzero.

In practice, this experiment is extremely difficult because of the enormous speed of light.

$$c = 3 \times 10^8 \text{ m/s}$$

If our speed relative to the ether is v , the observed speed of light should vary between $c - v$ and $c + v$. Although the value of v is unknown, it should on average be of the same order as the earth’s orbital velocity around the sun,

$$v \sim 3 \times 10^4 \text{ m/s}$$

(or possibly more if the sun is also moving relative to the ether frame). Thus the expected change in the observed speed of light due to the earth’s motion is

Section 1.4 • Classical Relativity and the Speed of Light 9

about 1 part in 10^4 . This was too small a change to be detected by direct measurement of the speed of light at that time.

To avoid the need for such direct measurements, Michelson devised an interferometer in which a beam of light was split into two beams by a partially reflecting surface; the two beams traveled along perpendicular paths and were then reunited to form an interference pattern; this pattern was sensitive to differences in the speed of light in the two perpendicular directions and so could be used to detect any such differences. By 1887, Michelson and Morley had built an interferometer (described below) that should have been able to detect differences in the speed of light much smaller than the part in 10^4 expected. To their surprise and chagrin, they could detect absolutely no difference at all.

The Michelson–Morley and similar experiments have been repeated many times, at different times of year and with ever-increasing precision, but always with the same final result.* With hindsight, it is easy to draw the right conclusion from their experiment: Contrary to all expectations, light always travels with the same speed in all directions relative to an earth-based reference frame even though the earth has different velocities at different times of the year. In other words, light travels at the same speed c in all directions in *many different* inertial frames, and the notion of a unique ether frame with this property must be abandoned.

This conclusion is so surprising that it was not taken seriously for nearly 20 years. Rather, several ingenious alternative theories were advanced that explained the Michelson–Morley result but managed to preserve the notion of a unique ether frame. For example, in the “ether-drag” theory, it was suggested that the ether, the medium through which light was supposed to propagate, was dragged along by the earth as it moved through space (in much the same way that the earth does drag its atmosphere with it). If this were the case, an earthbound observer would automatically be at rest relative to the ether, and Michelson and Morley would naturally have found that light had the same speed in all directions at all times of the year. Unfortunately, this neat explanation of the Michelson–Morley result requires that light from the stars would be bent as it entered the earth’s envelope of ether. Instead, astronomical observations show that light from any star continues to move in a straight line as it arrives at the earth.[†]

The ether-drag theory, like all other alternative explanations of the Michelson–Morley result, has been abandoned because it fails to fit all the facts. Today, nearly all physicists agree that Michelson and Morley’s failure to detect our motion relative to the ether frame was because there is no ether frame. The first person to accept this surprising conclusion and to develop its consequences into a complete theory was Einstein, as we describe, starting in Section 1.6.

Albert Michelson

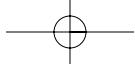
(1852–1931, American)



Michelson devoted much of his career to increasingly accurate measurements of the speed of light, and in 1907 he won the Nobel Prize in physics for his contributions to optics. His failure to detect the earth’s motion relative to the supposed ether is probably the most famous “unsuccessful” experiment in the history of science.

*From time to time experimenters have reported observing a nonzero difference, but closer examination has shown that these are probably due to spurious effects such as expansion and contraction of the interferometer arms resulting from temperature variations. For a careful modern analysis of Michelson and Morley’s results and many further references, see M. Handschy, *American Journal of Physics*, vol. 50, p. 987 (1982).

[†]Because of the earth’s motion around the sun, the apparent direction of any one star undergoes a slight annual variation—an effect called stellar aberration. This effect is consistent with the claim that light travels in a straight line from the star to the earth’s surface, but contradicts the ether-drag theory.



10 Chapter I • The Space and Time of Relativity

1.5 The Michelson–Morley Experiment*

*More than a hundred years later, the Michelson–Morley experiment remains the simplest and cleanest evidence that light travels at the same speed in all directions in all inertial frames — what became the second postulate of relativity. Naturally, we think you should know a little of how this historic experiment worked. Nevertheless, if you are pressed for time, you can omit this section without loss of continuity.

Figure 1.4 is a simplified diagram of Michelson’s interferometer. Light from the source hits the half-silvered mirror M and splits, part traveling to the mirror M_1 , and part to M_2 . The two beams are reflected at M_1 and M_2 and return to M , which sends part of each beam on to the observer. In this way the observer receives two signals, which can interfere constructively or destructively, depending on their phase difference.

To calculate this phase difference, suppose for a moment that the two arms of the interferometer, from M to M_1 and M to M_2 , have exactly the same length l , as shown. In this case any phase difference must be due to the different speeds of the two beams as they travel along the two arms. For simplicity, let us assume that arm 1 is exactly parallel to the earth’s velocity \mathbf{v} . In this case the light travels from M to M_1 with speed $c + v$ (relative to the interferometer) and back from M_1 to M with speed $c - v$. Thus the total time for the round trip on path 1 is

$$t_1 = \frac{l}{c+v} + \frac{l}{c-v} = \frac{2lc}{c^2-v^2} \quad (1.4)$$

It is convenient to rewrite this in terms of the ratio

$$\beta = \frac{v}{c}$$

which we have seen is expected to be very small, $\beta \sim 10^{-4}$. In terms of β , (1.4) becomes

$$t_1 = \frac{2l}{c} \frac{1}{1-\beta^2} \approx \frac{2l}{c} (1 + \beta^2) \quad (1.5)$$

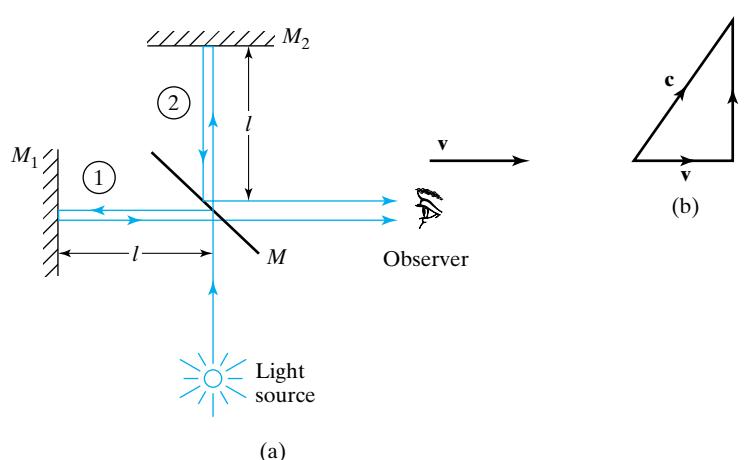


FIGURE 1.4

(a) Schematic diagram of the Michelson interferometer. M is a half-silvered mirror, M_1 and M_2 are mirrors. The vector \mathbf{v} indicates the earth’s velocity relative to the supposed ether frame. (b) The vector-addition diagram that gives the light’s velocity \mathbf{u} , relative to the earth, as it travels from M to M_2 . The velocity \mathbf{c} relative to the ether is the vector sum of \mathbf{v} and \mathbf{u} .

Section 1.5 • The Michelson–Morley Experiment



In the last step we have used the **binomial approximation** (discussed in Appendix B and in Problems 1.12–1.14),

$$(1 - x)^n \approx 1 - nx \quad (1.6)$$

which holds for any number n and any x much smaller than 1. (In the present case $n = -\frac{1}{2}$ and $x = \beta^2$.)

The speed of light traveling from M to M_2 is given by the velocity-addition diagram in Fig. 1.4(b). (Relative to the earth, the light has velocity \mathbf{u} perpendicular to \mathbf{v} ; relative to the ether, it travels with speed c in the direction shown.) This speed is

$$u = \sqrt{c^2 - v^2}$$

Since the speed is the same on the return journey, the total time for the round trip on path 2 is

$$t_2 = \frac{2l}{\sqrt{c^2 - v^2}} = \frac{2l}{c\sqrt{1 - \beta^2}} \approx \frac{2l}{c} \left(1 + \frac{1}{2}\beta^2\right) \quad (1.7)$$

where we have again used the binomial approximation (1.6), this time with $n = -\frac{1}{2}$.

Comparing (1.5) and (1.7), we see that the waves traveling along the two arms take slightly different times to return to M , the difference being

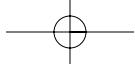
$$\Delta t = t_1 - t_2 \approx \frac{l}{c} \beta^2 \quad (1.8)$$

If this difference Δt were zero, the two waves would arrive in step and interfere constructively, giving a bright resultant signal. Similarly, if Δt were any integer multiple of the light's period, $T = \lambda/c$ (where λ is the wavelength), they would interfere constructively. If Δt were equal to half the period, $\Delta t = 0.5T$ (or $1.5T$, or $2.5T$, ...), the two waves would be exactly out of step and would interfere destructively. We can express these ideas more compactly if we consider the ratio

$$N = \frac{\Delta t}{T} = \frac{l\beta^2/c}{\lambda/c} = \frac{l\beta^2}{\lambda} \quad (1.9)$$

This is the number of complete cycles by which the two waves arrive out of step; in other words, N is the phase difference, expressed in cycles. If N is an integer, the waves interfere constructively; if N is a half-odd integer ($N = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$), the waves interfere destructively.

The phase difference N in (1.9) is the phase difference due to the earth's motion relative to the supposed ether frame. In practice, it is impossible to be sure that the two interferometer arms have exactly equal lengths, so there will be an additional phase difference due to the unknown difference in lengths. To circumvent this complication, Michelson and Morley rotated their interferometer through 90° , observing the interference as they did so. This rotation would not change the phase difference due to the different arm lengths, but it should reverse the phase difference due to the earth's motion (since arm 2 would now



12 Chapter I • The Space and Time of Relativity

be along \mathbf{v} and arm 1 across it). Thus, as a result of the rotation, the phase difference N should change by *twice* the amount (1.9),

$$\Delta N = \frac{2l\beta^2}{\lambda} \quad (1.10)$$

This implies that the observed interference should shift from bright to dark and back to bright again ΔN times. Observation of this shift would confirm that the earth is moving relative to the ether frame, and measurement of ΔN would give the value of β , and hence the earth's velocity $v = \beta c$.

In their experiment of 1887, Michelson and Morley had an arm length $l \approx 11$ m. (This was accomplished by having the light bounce back and forth between several mirrors.) The wavelength of their light was $\lambda = 590$ nm; and as we have seen, $\beta = v/c$ was expected to be of order 10^{-4} . Thus the shift should have been at least

$$\Delta N = \frac{2l\beta^2}{\lambda} \approx \frac{2 \times (11 \text{ m}) \times (10^{-4})^2}{590 \times 10^{-9} \text{ m}} \approx 0.4 \quad (1.11)$$

Although they could detect a shift as small as 0.01, Michelson and Morley observed no significant shift when they rotated their interferometer.

Michelson and Morley were disappointed and shocked at their result, and it was almost 20 years before anyone drew the right conclusion from it — that light has the same speed c in all directions in all inertial frames, the idea that Einstein adopted as one of the postulates of his theory of relativity.

Albert Einstein
(1879–1955,
German–Swiss–American)



Like all scientific theories, relativity was the work of many people. Nevertheless, Einstein's contributions outweigh those of anyone else by so much that the theory is quite properly regarded as his. As we will see in Chapter 4, he also made fundamental contributions to quantum theory, and it was for these that he was awarded the 1921 Nobel Prize in physics. The exotic ideas of relativity and the gentle, unpretentious persona of its creator excited the imagination of the press and public, and Einstein became the most famous scientist who ever lived. Asked what his profession was, the aged Einstein once answered, "photographer's model!"

1.6 The Postulates of Relativity

We have seen that the classical ideas of space and time had led to two conclusions:

1. The laws of Newtonian mechanics hold in an entire family of reference frames, any one of which moves uniformly relative to any other.
2. There can be only one reference frame in which light travels at the same speed c in all directions (and, more generally, in which all laws of electromagnetism are valid).

The Michelson–Morley experiment and numerous other experiments in the succeeding hundred years have shown that the second conclusion is false. Light travels with speed c in all directions in many different reference frames.

Einstein's special theory of relativity is based on the acceptance of this fact. Einstein proposed two postulates, or axioms, expressing his conviction that all physical laws, including mechanics *and* electromagnetism, should be valid in an entire family of reference frames. From these two postulates, he developed his special theory of relativity.

Before we state the two postulates of relativity, it is convenient to expand the definition of an **inertial frame** to be any reference frame in which *all* the laws of physics hold.

An inertial frame is any reference frame (that is, system of coordinates x , y , z and time t) where all the laws of physics hold in their simplest form.

Notice that we have not yet said what “all the laws of physics” are; to a large extent, Einstein used his postulates to *deduce* what the correct laws of physics could be. It turns out that one of the laws that survives from classical physics into relativity is Newton’s first law, the law of inertia. Thus our newly defined inertial frames are in fact the familiar unaccelerated frames where a body on which no forces act moves with constant velocity. As before, a reference frame anchored to the earth is an inertial frame (to the extent that we ignore the small accelerations due to the earth’s rotation and orbital motion); a reference frame fixed to a rapidly rotating turntable is not an inertial frame.

Notice also that in defining an inertial frame, we have specified that the laws of physics must hold “in their simplest form.” This is because one can sometimes modify physical laws so that they hold in noninertial frames as well. For example, by introducing a “fictitious” centrifugal force, one can arrange that the laws of statics are valid in a rotating frame. It is to exclude this kind of modification that we have added the qualification “in their simplest form.”

The first postulate of relativity asserts that there is a whole family of inertial frames.

FIRST POSTULATE OF RELATIVITY

If S is an inertial frame and if a second frame S' moves with constant velocity relative to S , then S' is also an inertial frame.

We can reword this postulate to say that the laws of physics are *invariant* as we change from one reference frame to a second frame, moving uniformly relative to the first. This property is familiar from classical mechanics, but in relativity it is postulated for all the laws of physics.

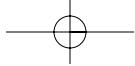
The first postulate is often paraphrased as follows: “There is no such thing as absolute motion.” To understand what this means, consider a frame S' attached to a rocket moving at constant velocity relative to a frame S anchored to the earth. The question we want to ask is this: Is there any scientific sense in which we can say that S' is really moving and that S is really stationary (or, perhaps, the other way around)? If the answer were “yes,” we could say that S is absolutely at rest and that anything moving relative to S is in absolute motion. However, the first postulate of relativity guarantees that this is impossible: All laws observable by an earthbound scientist in S are equally observable by a scientist in the rocket S' ; any experiment that can be performed in S can be performed equally in S' . Thus no experiment can possibly show which frame is “really” moving. Relative to the earth, the rocket is moving; relative to the rocket, the earth is moving; and this is as much as we can say.

Yet another way to express the first postulate is to say that among the family of inertial frames, all moving relative to one another, there is no *preferred frame*. That is, physics singles out no particular inertial frame as being in any way more special than any other frame.

The second postulate identifies one of the laws that holds in all inertial frames.

SECOND POSTULATE OF RELATIVITY

In all inertial frames, light travels through the vacuum with the same speed, $c = 299,792,458 \text{ m/s}$ in any direction.



14 Chapter I • The Space and Time of Relativity

This postulate is, of course, the formal expression of the Michelson–Morley result. We can say briefly that it asserts the *universality of the speed of light* c .

The second postulate flies in the face of our normal experience. Nevertheless, it is now a firmly established experimental fact. As we explore the consequences of the two postulates of relativity, we are going to encounter several unexpected effects that may be difficult to accept at first. All of these effects (including the second postulate itself) have the subtle property that they become important only when bodies travel at speeds reasonably close to the speed of light. Under ordinary conditions, at normal terrestrial speeds, these effects simply do not show up. In this sense, none of the surprising consequences of Einstein's relativity really contradicts our everyday experience.

1.7 Measurement of Time

Before we begin exploring the consequences of the relativity postulates, we need to say a word about the measurement of time. We are going to find that the time of an event may be different when measured from different frames of reference. This being the case, we must first be quite sure we know what we mean by measurement of time in a single frame.

It is implicit in the second postulate of relativity, with its reference to the speed of light, that we can measure distances and times. In particular, we take for granted that we have access to several accurate clocks. These clocks need not all be the same; but when they are all brought to the same point in the same inertial frame and are properly synchronized, they must of course agree.

Consider now a single inertial frame S , with origin O and axes x, y, z . We imagine an observer sitting at O and equipped with one of our clocks. Using her clock, the observer can easily time any event, such as a small explosion, in the immediate proximity of O since she will see (or hear) the event the moment it occurs. To time an event far away from O is harder, since the light (or sound) from the event has to travel to O before our observer can sense it. To avoid this complication, we let our observer hire a large number of helpers, each of whom she equips with an accurate clock and assigns to a fixed, known position in the coordinate system S , as shown in Fig. 1.5. Once the helpers are in position, she can check that their clocks are still synchronized by having each helper send a flash of light at an agreed time (measured on the helper's clock); since light travels with the known speed c (second postulate), she can calculate the time for the light to reach her at O and hence check the setting of the helper's clock.

With enough helpers, stationed closely enough together, we can be sure there is a helper sufficiently close to any event to time it effectively instantaneously. Once he has timed it, he can, at his leisure, inform everyone else of the result by any convenient means (by telephone, for example). In this way any event can be assigned a time t , as measured in the frame S .

When we speak of an inertial frame S , we will always have in mind a system of axes $Oxyz$ and a team of observers who are stationed at rest throughout S and equipped with synchronized clocks. This allows us to speak of the position $\mathbf{r} = (x, y, z)$ and the time t of any event, relative to the frame S .

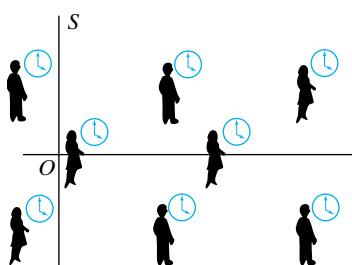


FIGURE 1.5

The chief observer at O distributes her helpers, each with an identical clock, throughout S .

1.8 The Relativity of Time; Time Dilation

We are now ready to compare measurements of times made by observers in two different inertial frames, and we are going to find that, as a consequence of the relativity postulates, times measured in different frames inevitably disagree. To this end, we imagine the familiar two frames, S anchored to the ground and S' anchored to a train moving at constant velocity v relative to the ground. We consider a “thought experiment” (or “gedanken experiment” from the German) in which an observer at rest on the train sets off a flashbulb on the floor of the train, vertically below a mirror mounted on the roof, a height h above. As seen in the frame S' (fixed in the train), a pulse of light travels straight up to the mirror, is reflected straight back, and returns to its starting point on the floor. We can imagine a photocell arranged to give an audible beep as the light returns. Our object is to find the time, as measured in either frame, between the two events — the flash as the light leaves the floor and the beep as it returns.

Our experiment, as seen in the frame S' , is shown in Fig. 1.6(a). Since S' is an inertial frame, light travels the total distance $2h$ at speed c . Therefore, the time for the entire trip is

$$\Delta t' = \frac{2h}{c} \quad (1.12)$$

This is the time that an observer in frame S' will measure between the flash and the beep, provided of course, that his clock is reliable.

The same experiment, as seen from the inertial frame S , is shown in Fig. 1.6(b). In this frame the light travels along the two sides AB and BC of the triangle shown. If we denote by Δt the time for the entire journey, as measured in S , the time to go from A to B is $\Delta t/2$. During this time the train travels a distance $v\Delta t/2$, and the light, moving with speed c , travels a distance $c\Delta t/2$. (Note that this is where the postulates of relativity come in; we have taken the speed of light to be c in both S and S' .) The dimensions of the right triangle

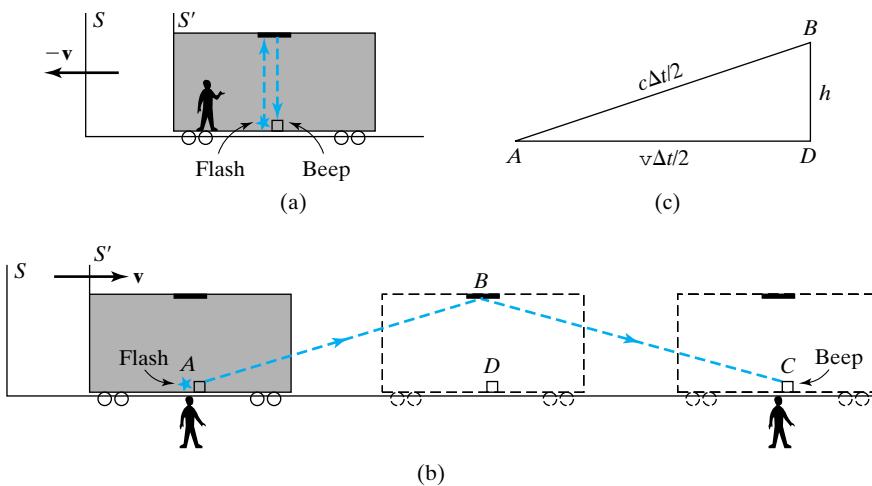
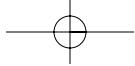


FIGURE 1.6

- (a) The thought experiment as seen in the train-based frame S' .
- (b) The same experiment as seen from the ground-based frame S . Notice that two observers are needed in this frame.
- (c) The dimensions of the triangle ABD .



16 Chapter I • The Space and Time of Relativity

ABD are therefore as shown in Fig. 1.6(c). Applying Pythagoras's theorem, we see that*

$$\left(\frac{c \Delta t}{2}\right)^2 = h^2 + \left(\frac{v \Delta t}{2}\right)^2$$

or, solving for Δt ,

$$\Delta t = \frac{2h}{\sqrt{c^2 - v^2}} = \frac{2h}{c} \frac{1}{\sqrt{1 - \beta^2}} \quad (1.13)$$

where we have again used the ratio

$$\beta = \frac{v}{c}$$

of the speed v to the speed of light c . The time Δt is the time that observers in S will measure between the flash and the beep (provided, again, that their clocks are reliable).

The most important and surprising thing about the two answers (1.12) and (1.13) is that they are not the same. The time between the two events, the flash and the beep, is different as measured in the frames S and S' . Specifically,

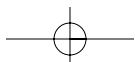
$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \beta^2}} \quad (1.14)$$

We have derived this result for an imagined thought experiment involving a flash of light reflected back to a photocell. However, the conclusion applies to any two events that occur at the same place on the train: Suppose, for instance, that we drop a knife on the table and a moment later drop a fork. In principle, at least, we could arrange for a flash of light to occur at the moment the knife lands, and we could position a mirror to reflect the light back to arrive just as the fork lands. The relation (1.14) must then apply to these two events (the landing of the knife and the landing of the fork). Now the falling of the knife and fork cannot be affected by the presence or absence of a flashbulb and photocell; thus neither of the times Δt or $\Delta t'$ can depend on whether we actually did the experiment with the light and the photocell. Therefore, the relation (1.14) holds for any two events that occur at the same place on board the train.

The difference between the measured times Δt and $\Delta t'$ is a direct consequence of the second postulate of relativity. (In classical physics $\Delta t = \Delta t'$, of course.) You should avoid thinking that the clocks in one of our frames must somehow be running wrong; quite the contrary, it was an essential part of our argument that all the clocks were running right. Moreover, our argument made no reference to the kind of clocks used (apart from requiring that they be correct). Thus the difference (1.14) applies to all clocks. In other words, time itself as measured in the two frames is different. We will discuss the experimental evidence for this surprising conclusion shortly.

Several properties of the relationship (1.14) deserve comment. First, if our train is actually at rest ($v = 0$), then $\beta = 0$ and (1.14) tells us that $\Delta t = \Delta t'$. That is, there is no difference unless the two frames are in relative

*Here we are taking for granted that the height h of the train is the same as measured in either frame, S or S' . We will prove that this is correct in Section 1.10.



Section 1.8 • The Relativity of Time; Time Dilation 17

motion. Further, at normal terrestrial speeds, $v \ll c$ and $\beta \ll 1$; thus the difference between Δt and $\Delta t'$ is very small.

Example 1.1

The pilot of a jet traveling at a steady 300 m/s sets a buzzer in the cockpit to go off at intervals of exactly 1 hour (as measured on the plane). What would be the interval between two successive buzzes as measured by two observers suitably positioned on the ground? (Ignore effects of the earth's motion; that is, consider the ground to be an inertial frame.)

The required interval between two buzzes is given by (1.14), with $\Delta t' = 1$ hour and $\beta = v/c = 10^{-6}$. Thus

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \beta^2}} = \frac{1 \text{ hour}}{\sqrt{1 - 10^{-12}}}$$

We have to be a bit careful in evaluating this time. The number in the denominator is so close to 1 that most calculators cannot tell the difference. (It takes 12 significant figures to distinguish $1 - 10^{-12}$ from 1.) In this situation the simplest and best course is to use the binomial approximation, $(1 - x)^n \approx 1 - nx$, which is an excellent approximation, provided x is small. [This important approximation was already used in (1.6) and is discussed in Problems 1.12–1.14 and in Appendix B.] In the present case, setting $x = \beta^2$ and $n = -1/2$, we find

$$\begin{aligned}\Delta t &= \Delta t'(1 - \beta^2)^{-1/2} \approx \Delta t'\left(1 + \frac{1}{2}\beta^2\right) \\ &= (1 \text{ hour}) \times \left(1 + \frac{1}{2} \times 10^{-12}\right) \\ &= 1.000000000005 \text{ hours}\end{aligned}$$

The difference between the two measured times is 5×10^{-13} hour, or 1.8 nanoseconds. (A nanosecond, or ns, is 10^{-9} s.) It is easy to see why classical physicists had failed to notice this kind of difference!

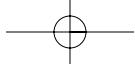
The difference between Δt and $\Delta t'$ gets bigger as v increases. In modern particle accelerators it is common to have electrons and other particles with speeds of $0.99c$ and more. If we imagine repeating our thought experiment with the frame S' attached to an electron with $\beta = 0.99$, Eq. (1.14) gives

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - (0.99)^2}} \approx 7\Delta t'$$

Differences as large as this are routinely observed by particle physicists, as we discuss in the next section.

If we were to put $v = c$ (that is, $\beta = 1$) in Eq. (1.14), we would get the absurd result, $\Delta t = \Delta t'/0$; and if we put $v > c$ (that is, $\beta > 1$), we would get an imaginary answer. These ridiculous results suggest (correctly) that v must always be less than c .

$$v < c$$



18 Chapter I • The Space and Time of Relativity

This is one of the most profound results of Einstein's relativity: *The speed of any inertial frame relative to any other inertial frame must always be less than c .* In other words, the speed of light, in addition to being the same in all inertial frames, emerges as the universal speed limit for the relative motion of inertial frames.

The factor $1/\sqrt{1 - \beta^2}$ that appears in Eq. (1.14) crops up in so many relativistic formulas that it is traditionally given its own symbol, γ .

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} = \frac{1}{\sqrt{1 - (v/c)^2}} \quad (1.15)$$

Since v is always smaller than c , the denominator in (1.15) is always less than or equal to 1 and hence

$$\gamma \geq 1 \quad (1.16)$$

The factor γ equals 1 only if $v = 0$. The larger we make v , the larger γ becomes; and as v approaches c , the value of γ increases without limit.

In terms of γ , Eq. (1.14) can be rewritten

$$\Delta t = \gamma \Delta t' \geq \Delta t' \quad (1.17)$$

That is, Δt is always greater than or equal to $\Delta t'$. This asymmetry may seem surprising, and even to violate the postulates of relativity since it suggests a special role for the frame S' . In fact, however, this is just as it should be. In our experiment the frame S' is special since it is the unique inertial frame where the two events — the flash and the beep — occurred *at the same place*. This asymmetry was implicit in Fig. 1.6, which showed *one* observer measuring $\Delta t'$ (since both events occurred at the same place in S') but *two* observers measuring Δt (since the two events were at different places in S). To emphasize this asymmetry, the time $\Delta t'$ can be renamed Δt_0 and (1.17) rewritten as

$$\Delta t = \gamma \Delta t_0 \geq \Delta t_0 \quad (1.18)$$

The subscript 0 on Δt_0 indicates that Δt_0 is the time indicated by a clock that is at rest in the special frame where the two events occurred at the same place. This time is often called the **proper time** between the events. The time Δt is measured in *any* frame and is always greater than or equal to the proper time Δt_0 . For this reason, the effect embodied in (1.18) is often called **time dilation**.

The proper time Δt_0 is the time indicated by the clock on the moving train (moving relative to S , that is); Δt is the time shown by the clocks at rest on the ground in frame S . Since $\Delta t_0 \leq \Delta t$, the relation (1.18) can be loosely paraphrased to say that “*a moving clock is observed to run slow*.”

Finally, we should reemphasize the fundamental symmetry between any two inertial frames. We chose to conduct our thought experiment with the flash and beep at one spot on the train (frame S'), and we found that $\Delta t > \Delta t'$. However, we could have done things the other way around: If a ground-based observer (at rest in S) had performed the same experiment with a flash of light and a mirror, the flash and beep would have occurred in the same spot on the ground; and we would have found that $\Delta t' \geq \Delta t$. The great merit of writing the time-dilation formula in the form (1.18), $\Delta t = \gamma \Delta t_0$, is

that it avoids the problem of remembering which is frame S and which S' ; the subscript 0 always identifies the proper time, as measured in the frame in which the two events were at the same spot.

1.9 Evidence for Time Dilation

In his original paper on relativity, Einstein predicted the effect that is now called time dilation. At that time there was no evidence to support the prediction, and many years were to pass before any was forthcoming. The first tests, using the unstable particle called the muon as their “clock,” were carried out in 1941. (See Problem 1.27.)

It was only with the advent of super-accurate atomic clocks that tests using man-made clocks became possible. The first such test was carried out in 1971. Four portable atomic clocks were synchronized with a reference clock at the U.S. Naval Observatory in Washington, D.C., and all four clocks were then flown around the world on a jet plane and returned to the Naval Observatory. The discrepancy between the reference clock and the portable clocks after their journey was predicted (using relativity) to be

$$275 \pm 21 \text{ ns} \quad (1.19)$$

while the observed discrepancy (averaged over the four portable clocks) was*

$$273 \pm 7 \text{ ns} \quad (1.20)$$

We should mention that the excellent agreement between (1.19) and (1.20) is more than a test of the time difference (1.18), predicted by special relativity. Gravitational effects, which require general relativity, contribute a large part of the predicted discrepancy (1.19). Thus this beautiful experiment is a confirmation of general, as well as special, relativity.

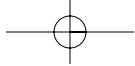
Much simpler tests of time dilation and tests involving much larger dilations are possible if one is prepared to use the natural clocks provided by unstable subatomic particles. For example, the charged π meson, or pion, is a particle that is formed in collisions between rapidly moving atomic nuclei (as we discuss in detail in Chapter 18). The pion has a definite average lifetime, after which it “decays” or disintegrates into other subatomic particles, and one can use this average life as a kind of natural clock.

One way to characterize the life span of an unstable particle is the **half-life**[†] $t_{1/2}$, the average time after which half of a large sample of the particles in question will have decayed. For example, the half-life of the pion is measured to be

$$t_{1/2} = 1.8 \times 10^{-8} \text{ s} \quad (1.21)$$

*The test was actually carried out twice — once flying east and once west — with satisfactory agreement in both cases. The results quoted here are from the more decisive westward flight. For more details, see J. C. Hafele and R. E. Keating, *Science*, vol. 177, p. 166 (1972). Since the accuracy of this original experiment has been questioned, we should emphasize that the experiment has been repeated many times, with improved accuracy, and there is now no doubt at all that the observations support the predictions of relativity.

[†]An alternative characterization is the mean life τ , which differs from $t_{1/2}$ by a constant factor. We will define both of these more carefully in Chapter 17.



20 Chapter I • The Space and Time of Relativity

This means that if one starts at $t = 0$ with N_0 pions, then after 1.8×10^{-8} s half of them will have decayed and only $N_0/2$ will remain. After a further 1.8×10^{-8} s, half of those $N_0/2$ will have decayed and only $N_0/4$ will remain. After another 1.8×10^{-8} s, only $N_0/8$ will remain. And so on. In general, after n half-lives, $t = n t_{1/2}$, the number of particles remaining will be $N_0/2^n$.

At particle-physics laboratories, pions are produced in large numbers in collisions between protons (the nuclei of hydrogen atoms) and various other nuclei. It is usually convenient to conduct experiments with the pions at a good distance from where they are produced, and the pions are therefore allowed to fly down an evacuated pipe to the experimental area. At the Fermilab near Chicago the pions are produced traveling very close to the speed of light, a typical value being

$$v = 0.9999995c$$

and the distance they must travel to the experimental area is about $L = 1$ km. Let us consider the flight of these pions, first from the (incorrect) classical view with no time dilation and then from the (correct) relativistic view.

As seen in the laboratory, the pions' time of flight is

$$T = \frac{L}{v} \approx \frac{10^3 \text{ m}}{3 \times 10^8 \text{ m/s}} = 3.3 \times 10^{-6} \text{ s} \quad (1.22)$$

A classical physicist, untroubled by any notions of relativity of time, would compare this with the half-life (1.21) and calculate that

$$T \approx 183t_{1/2}$$

That is, the time needed for the pions to reach the experimental area is 183 half-lives. Therefore, if N_0 is the original number of pions, the number to survive the journey would be

$$N = \frac{N_0}{2^{183}} \approx (8.2 \times 10^{-56})N_0$$

and for all practical purposes, *no* pions would reach the experimental area. This would obviously be an absurd way to do experiments with pions, and it is not what actually happens.

In relativity, we now know, times depend on the frame in which they are measured, and we must consider carefully the frames to which the times T and $t_{1/2}$ refer. The time T in (1.22) is, of course, the time of flight of the pions as measured in a frame fixed in the laboratory, the *lab frame*. To emphasize this, we rewrite (1.22) as

$$T(\text{lab frame}) = 3.3 \times 10^{-6} \text{ s} \quad (1.23)$$

On the other hand, the half-life $t_{1/2} = 1.8 \times 10^{-8}$ s refers to time as “seen” by the pions; that is, $t_{1/2}$ is the half-life measured in a frame anchored to the pions, the pions’ **rest frame**. (This is an experimental fact: The half-lives quoted by physicists are the proper half-lives, measured in the frame where the particles are at rest.) To emphasize this, we write (temporarily)

$$t_{1/2}(\pi \text{ rest frame}) = 1.8 \times 10^{-8} \text{ s} \quad (1.24)$$

We see that the classical argument here used two times, T and $t_{1/2}$ measured in *different* inertial frames. A correct argument must work consistently in one frame, for example the lab frame. The half-life measured in the lab frame is given by the time-dilation formula as γ times the half-life (1.24). With $\beta = 0.9999995$, it is easy to see that

$$\gamma = 1000$$

and hence that

$$\begin{aligned} t_{1/2}(\text{lab frame}) &= \gamma t_{1/2}(\pi \text{ rest frame}) \\ &= 1000 \times (1.8 \times 10^{-8} \text{ s}) \\ &= 1.8 \times 10^{-5} \text{ s} \end{aligned} \quad (1.25)$$

Comparing (1.23) and (1.25) we see that

$$T(\text{lab frame}) \approx 0.2t_{1/2}(\text{lab frame})$$

That is, the pions' flight down the pipe lasts only one-fifth of the relevant half-life. In this time very few of the pions decay, and almost all reach the experimental area. (The number that survive is $N = N_0/2^{0.2} \approx 0.9N_0$.) That this is exactly what actually happens in all particle-physics laboratories is powerful confirmation of the relativity of time, as first predicted by Einstein in 1905.

Example 1.2

The Λ particle (Λ is the Greek capital L and is pronounced “lambda.”) is an unstable subatomic particle that decays into a proton and a pion ($\Lambda \rightarrow p + \pi$) with a half-life of $t_{1/2} = 1.7 \times 10^{-10} \text{ s}$. If several lambdas are created in a nuclear collision, all with speed $v = 0.6 c$, on average how far will they travel before half of them decay?

The half-life as measured in the laboratory is $\gamma t_{1/2}$ (since $t_{1/2}$ is the proper half-life, as measured in the Λ rest frame). Therefore, the desired distance is $v\gamma t_{1/2}$. With $\beta = 0.6$,

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} = 1.25$$

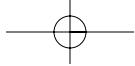
and the required distance is

$$\text{distance} = v\gamma t_{1/2} = (1.8 \times 10^8 \text{ m/s}) \times 1.25 \times (1.7 \times 10^{-10} \text{ s}) = 3.8 \text{ cm}$$

Notice how even with speeds as large as $0.6 c$, the factor γ is not very much larger than 1, and the effect of time dilation is not dramatic. Notice also that a distance of a few centimeters is much easier to measure than a time of order 10^{-10} s ; thus measurement of the range of an unstable particle is often the easiest way to find its half-life.

1.10 Length Contraction

The postulates of relativity have led us to conclude that time depends on the reference frame in which it is measured. We can now use this fact to show that the same must also apply to distances: The measured distance between two



22 Chapter I • The Space and Time of Relativity

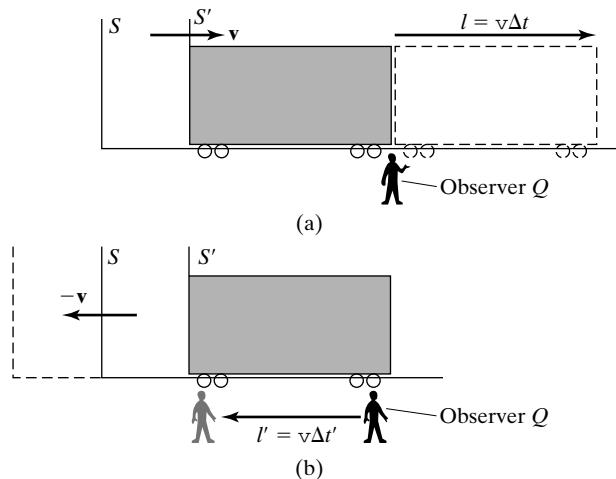


FIGURE 1.7

(a) As seen in S , the train moves a distance $v \Delta t$ to the right. (b) As seen in S' , the frame S and observer Q move a distance $v \Delta t'$ to the left.

events depends on the frame relative to which it is measured. We will show this with another thought experiment. In the analysis of this thought experiment, it will be important to recognize that, even in relativity, the familiar kinematic relation

$$\text{distance} = \text{velocity} \times \text{time}$$

is valid in any given inertial frame (with all quantities measured in that frame), since it is just the definition of velocity in that frame.

We imagine again our two frames, S fixed to the ground and S' fixed to a train traveling at velocity v relative to the ground; and we now imagine observers in S and S' measuring the length of the train. For an observer in S' , this measurement is easy since he sees the train at rest and can take all the time he needs to measure the length l' with an accurate ruler. For an observer Q on the ground, the measurement is harder since the train is moving. Perhaps the simplest procedure is to time the train as it passes Q [Fig. 1.7(a)]. If t_1 and t_2 are the times at which the front and back of the train pass Q and if $\Delta t = t_2 - t_1$, then Q can calculate the length l (measured in S) as

$$l = v \Delta t \quad (1.26)$$

To compare this answer with l' , we note that observers on the train could have measured l' by a similar procedure. As seen from the train, the observer Q on the ground is moving to the left with speed* v , and observers on the train can measure the time for Q to move from the front to the back of the train as in Fig. 1.7(b). (This would require two observers on the train, one at the front and one at the back.) If this time is $\Delta t'$,

$$l' = v \Delta t' \quad (1.27)$$

Comparing (1.26) and (1.27), we see immediately that since the times Δt and $\Delta t'$ are different, the same must be true of the lengths l and l' . To calculate

*We are taking for granted that the speed of S relative to S' is the same as that of S' relative to S . This follows from the basic symmetry between S and S' as required by the postulates of relativity.

the difference, we need to relate Δt and $\Delta t'$ using the time-dilation formula. In the present experiment the two events of interest, “ Q opposite the train’s front” and “ Q opposite the train’s back,” occur at the same place in S (where Q is at rest). Therefore, the time-dilation formula implies that

$$\Delta t' = \gamma \Delta t$$

Comparing (1.26) and (1.27), we see that

$$l = \frac{l'}{\gamma} \leq l' \quad (1.28)$$

The length of the train as measured in S is less than (or equal to) that measured in S' .

Like time dilation, this result is asymmetric, reflecting the asymmetry of our experiment: The frame S' is special since it is the unique frame where the measured object (the train) is at rest. [We could, of course, have done the experiment the other way around; if we had measured the length of a house at rest in S , the roles of l and l' in (1.28) would have been reversed.] To emphasize this asymmetry and to avoid confusion as to which frame is which, it is a good idea to rewrite (1.28) as

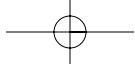
$$l = \frac{l_0}{\gamma} \leq l_0 \quad (1.29)$$

where the subscript 0 indicates that l_0 is the length of an object *measured in its rest frame*, while l refers to the length measured in *any frame*. The length l_0 can be called the object’s **proper length**. Since $l \leq l_0$, the effect implied by (1.29) is often called **length contraction** (or Lorentz contraction, or Lorentz–Fitzgerald contraction, after the two physicists who first suggested some such effect). The effect can be loosely described by saying that a *moving object is observed to be contracted*.

Evidence for Length Contraction

Like time dilation, length contraction is a real effect that is well established experimentally. Perhaps the simplest evidence comes from the same experiment as that discussed in connection with time dilation, in which unstable pions fly down a pipe from the collision that produces them to the experimental area. As viewed from the lab frame, we saw that time dilation increases the pions’ half-life by a factor of γ , from $t_{1/2}$ to $\gamma t_{1/2}$. In the example discussed, it was this increase that allowed most of the pions to complete the journey to the experimental area before they decayed.

Suppose, however, that we viewed the same experiment from the pions’ rest frame. In this frame the pions are stationary and there is no time dilation to increase their half-life. So how do they reach the experimental area? The answer is that in this frame the *pipe* is moving, and length contraction reduces its length by the same factor γ , from L to L/γ . Thus observers in this frame would say it is length contraction that allows the pions to reach the experimental area. Naturally, the number of pions completing the journey is the same whichever frame we use for the calculation.



24 Chapter I • The Space and Time of Relativity

Example I.3

A space explorer of a future era travels to the nearest star, Alpha Centauri, in a rocket with speed $v = 0.9c$. The distance from earth to the star, as measured from earth, is $L = 4$ light years. What is this distance as measured by the explorer, and how long will she say the journey to the star lasts? [A light year is the distance traveled by light in one year, which is just c multiplied by 1 year, or 9.46×10^{12} kilometers. In many problems it is better to write it as $1 c \cdot \text{year}$, since the c often cancels out, as we will see.]

The distance $L = 4 c \cdot \text{years}$ is the proper distance between earth and the star (which we assume are relatively at rest). Thus the distance as seen from the rocket is given by the length-contraction formula as

$$L(\text{rocket frame}) = \frac{L(\text{earth frame})}{\gamma}$$

If $\beta = 0.9$, then $\gamma = 2.3$, so

$$L(\text{rocket frame}) = \frac{4 c \cdot \text{years}}{2.3} = 1.7 c \cdot \text{years}$$

We can calculate the time T for the journey in two ways: As seen from the rocket, the star is initially $1.7 c \cdot \text{years}$ away and is approaching with speed $v = 0.9c$. Therefore,

$$\begin{aligned} T(\text{rocket frame}) &= \frac{L(\text{rocket frame})}{v} \\ &= \frac{1.7 c \cdot \text{years}}{0.9c} = 1.9 \text{ years} \end{aligned} \quad (1.30)$$

(Notice how the factors of c conveniently cancel when we use $c \cdot \text{years}$ and measure speeds as multiples of c .)

Alternatively, as measured from the earth frame, the journey lasts for a time

$$T(\text{earth frame}) = \frac{L(\text{earth frame})}{v} = \frac{4 c \cdot \text{years}}{0.9c} = 4.4 \text{ years}$$

But because of time dilation, this is γ times $T(\text{rocket frame})$, which is therefore

$$T(\text{rocket frame}) = \frac{T(\text{earth frame})}{\gamma} = 1.9 \text{ years}$$

in agreement with (1.30), of course.

Notice how time dilation (or length contraction) allows an appreciable saving to the pilot of the rocket. If she returns promptly to earth, then as a result of the complete round trip she will have aged only 3.8 years, while her twin who stayed behind will have aged 8.8 years. This surprising result, sometimes known as the **twin paradox**, is amply verified by the experiments discussed in Section 1.9. In principle, time dilation would allow explorers to

make in one lifetime trips that would require hundreds of years as viewed from earth. Since this requires rockets that travel very close to the speed of light, it is not likely to happen soon! See Problem 1.22 for further discussion of this effect.

Lengths Perpendicular to the Relative Motion

We have so far discussed lengths that are parallel to the relative velocity, such as the length of a train in its direction of motion. What happens to lengths perpendicular to the relative velocity, such as the height of the train? It is fairly easy to show that for such lengths, there is no contraction or expansion. To see this, consider two observers, Q at rest in S and Q' at rest in S' , and suppose that Q and Q' are equally tall when at rest. Now, let us assume for a moment that there is a contraction of heights analogous to the length contraction (1.29). If this is so, then as seen by Q , Q' will be shorter as he rushes by. We can test this hypothesis by having Q' hold up a sharp knife exactly level with the top of his head; if Q' is shorter, Q will find himself scalped (or worse) as the knife goes by.

This experiment is completely symmetric between the two frames S and S' : There is one observer at rest in each frame, and the only difference is the direction in which each sees the other moving.* Therefore, it must also be true that as seen by Q' , it is Q who is shorter. But this implies that the knife will miss Q . Since it cannot be true that Q is both scalped and not scalped, we have arrived at a contradiction, and there can be no contraction. By a similar argument, there can be no expansion, and, in fact, the knife held by Q' simply grazes past Q 's scalp, as seen in either frame. We conclude that lengths perpendicular to the relative motion are unchanged; and the Lorentz-contraction formula (1.29) applies only to lengths parallel to the relative motion.

1.11 The Lorentz Transformation

We are now ready to answer an important general question: If we know the coordinates x , y , z and time t of an event, as measured in a frame S , how can we find the coordinates x' , y' , z' , and t' of the same event as measured in a second frame S' ? Before we derive the correct relativistic answer to this question, we examine briefly the classical answer.

We consider our usual two frames, S anchored to the ground and S' anchored to a train traveling with velocity \mathbf{v} relative to S , as shown in Fig. 1.8. Because the laws of physics are all independent of our choice of origin and orientation, we are free to choose both axes Ox and $O'x'$ along the same line, parallel to \mathbf{v} , as shown. We can further choose the origins of time so that $t = t' = 0$ at the moment when O' passes O . We will sometimes refer to this arrangement of systems S and S' as the **standard configuration**.

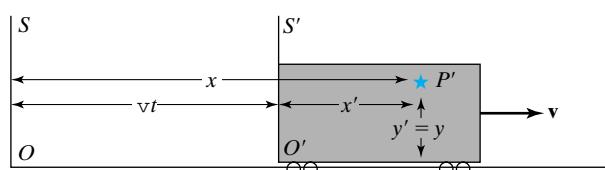
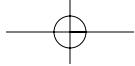


FIGURE 1.8

In classical physics the coordinates of an event are related as shown.

*Note that our previous two thought experiments were asymmetric, requiring two observers in one of the frames, but only one in the other.



26 Chapter I • The Space and Time of Relativity

Galileo Galilei
(1564–1642, Italian)



Considered by some the father of modern science, Galileo understood the importance of experiment and theory and was a master of both. Although he did not discover the telescope, he improved it and was the first to use it as a tool of astronomy, discovering the mountains on the moon, phases of Venus, moons of Jupiter, stars of the Milky Way, and sunspots and rotation of the sun. Among his many contributions to mechanics, he established the law of inertia and proved that gravity accelerates all bodies equally and that the period of a small-amplitude pendulum is independent of the amplitude. He understood clearly that the laws of mechanics hold in all unaccelerated frames, arguing that inside an enclosed cabin it would be impossible to detect the uniform motion of a ship. This argument appeared in his *Dialogue on the Two Chief World Systems* and was used to show that the earth could perfectly well be moving in orbit around the sun without our being aware of it in everyday life. For publishing this book, he was found guilty of heresy by the Holy Office of the Inquisition, and his book was placed on the Index of Prohibited Books — from which it was not removed until 1835.

Now consider an event, such as the explosion of a small firecracker, that occurs at position x, y, z , and time t as measured in S . Our problem is to calculate, in terms of x, y, z, t , (and the velocity v) the coordinates x', y', z', t' of the same event, as measured in S' — accepting at first the classical ideas of space and time. First, since time is a universal quantity in classical physics, we know that $t' = t$. Next, from Fig. 1.8 it is easily seen that $x' = x - vt$ and $y' = y$ (and, similarly, $z' = z$ although the z coordinate is not shown in the figure). Thus, according to the ideas of classical physics,

$$\begin{aligned}x' &= x - vt \\y' &= y \\z' &= z \\t' &= t\end{aligned}\quad (1.31)$$

These four equations are often called the **Galilean transformation** after Galileo Galilei, who was the first person known to have considered the invariance of the laws of motion under this change of coordinates. They transform the coordinates x, y, z, t of any event as observed in S into the corresponding coordinates x', y', z', t' as observed in S' .

If we had been given the coordinates x', y', z', t' and wanted to find x, y, z, t , we could solve the equations (1.31) to give

$$\begin{aligned}x &= x' + vt' \\y &= y' \\z &= z' \\t &= t'\end{aligned}\quad (1.32)$$

Notice that the equations (1.32) can be obtained directly from (1.31) by exchanging x, y, z, t with x', y', z', t' and replacing v by $-v$. This is because the relation of S to S' is the same as that of S' to S except for a change in the sign of the relative velocity.

The Galilean transformation (1.31) cannot be the correct relativistic relation between x, y, z, t , and x', y', z', t' . (For instance, we know from time dilation that the equation $t' = t$ cannot possibly be correct.) On the other hand, the Galilean transformation agrees perfectly with our everyday experience and so must be correct (to an excellent approximation) when the speed v is small compared to c . Thus the correct relation between x, y, z, t and x', y', z', t' will have to reduce to the Galilean relation (1.31) when v/c is small.

To find the correct relation between x, y, z, t and x', y', z', t' , we consider the same experiment as before, which is shown again in Fig. 1.9. We have noted before that distances perpendicular to \mathbf{v} are the same whether measured in S or S' . Thus

$$y' = y \quad \text{and} \quad z' = z \quad (1.33)$$

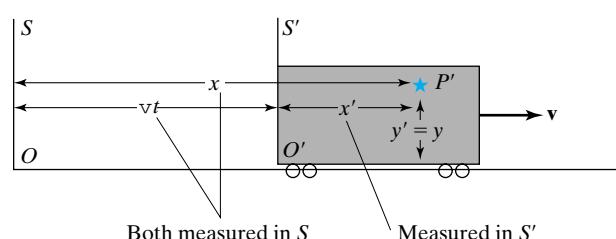


FIGURE 1.9

The coordinate x' is measured in S' . The distances x and vt are measured at the same time t in the frame S .

Section 1.11 • The Lorentz Transformation **27**

exactly as in the Galilean transformation. In finding x' , it is important to keep careful track of the frames in which the various quantities are measured; in addition, it is helpful to arrange that the explosion whose coordinates we are discussing produces a small burn mark on the wall of the train at the point P' where it occurs. The horizontal distance from the origin O' to the mark at P' , as measured in S' , is precisely the desired coordinate x' . Meanwhile, the same distance, as measured in S , is $x - vt$ (since x and vt are the horizontal distances from O to P' and O to O' at the instant t , as measured in S). Thus according to the length-contraction formula (1.29),

$$x - vt = \frac{x'}{\gamma}$$

or

$$x' = \gamma(x - vt) \quad (1.34)$$

This gives x' in terms of x and t and is the third of our four required equations. Notice that if v is small, $\gamma \approx 1$ and the relation (1.34) reduces to the first of the Galilean relations (1.31), as required.

Finally, to find t' in terms of x , y , z , and t , we use a simple trick. We can repeat the argument leading to (1.34) but with the roles of S and S' reversed. That is, we let the explosion burn a mark at the point P on a wall fixed in S , and arguing as before, we find that

$$x = \gamma(x' + vt) \quad (1.35)$$

[This can be obtained directly from (1.34) by exchanging x , t with x' , t' and replacing v by $-v$.] Equation (1.35) is not yet the desired result, but we can combine it with (1.34) to eliminate x' and find t' . Inserting (1.34) in (1.35), we get

$$x = \gamma[\gamma(x - vt) + vt']$$

Solving for t' we find that

$$t' = \gamma t - \frac{\gamma^2 - 1}{\gamma v} x$$

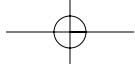
or, after some algebra (Problem 1.37),

$$t' = \gamma \left(t - \frac{vx}{c^2} \right) \quad (1.36)$$

This is the required expression for t' in terms of x and t . When v/c is much smaller than 1, we can neglect the second term, and since $\gamma \approx 1$, we get $t' \approx t$, in agreement with the Galilean transformation, as required.

Collecting together (1.33), (1.34), and (1.36), we obtain our required four equations.

$x' = \gamma(x - vt)$ $y' = y$ $z' = z$ $t' = \gamma \left(t - \frac{vx}{c^2} \right)$	(1.37)
---	--------



28 Chapter I • The Space and Time of Relativity

Hendrik Lorentz
(1853–1928, Dutch)



Lorentz was the first to write down the equations we now call the Lorentz transformation, although Einstein was the first to interpret them correctly. He also preceded Einstein with the length contraction formula (though, again, he did not interpret it correctly). He was one of the first to suggest that electrons are present in atoms, and his theory of electrons earned him the 1902 Nobel Prize in physics.

These equations are called the **Lorentz transformation**, or **Lorentz–Einstein transformation**, in honor of the Dutch physicist Lorentz, who first proposed them, and Einstein, who first interpreted them correctly. The Lorentz transformation is the correct relativistic modification of the Galilean transformation (1.31).

If one wants to know x, y, z, t in terms of x', y', z', t' , one can simply exchange the primed and unprimed variables and replace v by $-v$, in the now familiar way, to give

$$\begin{aligned}x &= \gamma(x' + vt') \\y &= y' \\z &= z' \\t &= \gamma\left(t' + \frac{vx'}{c^2}\right)\end{aligned}\quad (1.38)$$

These equations are sometimes called the **inverse Lorentz transformation**.

The Lorentz transformation expresses all the properties of space and time that follow from the postulates of relativity. From it, one can calculate all of the kinematic relations between measurements made in different inertial frames. In the next two sections we give some examples of such calculations.

1.12 Applications of the Lorentz Transformation

In this section we give three examples of problems that can easily be analyzed using the Lorentz transformation. In the first two we rederive two familiar results; in the third we analyze one of the many apparent paradoxes of relativity.

Example 1.4

Starting with the equations (1.37) of the Lorentz transformation, derive the length-contraction formula (1.29).

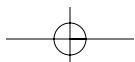
Notice that the length-contraction formula was used in our derivation of the Lorentz transformation. Thus this example will not give a new proof of length contraction; it will, rather, be a consistency check on the Lorentz transformation, to verify that it gives back the result from which it was derived.

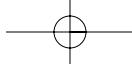
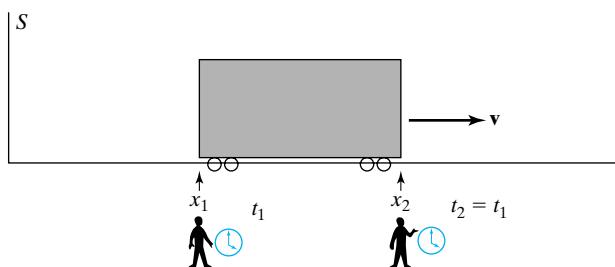
Let us imagine, as before, measuring the length of a train (frame S') traveling at speed v relative to the ground (frame S). If the coordinates of the back and front of the train are x'_1 and x'_2 , as measured in S' , the train's proper length (its length as measured in its rest frame) is

$$l_0 = l' = x'_2 - x'_1 \quad (1.39)$$

To find the length l as measured in S , we carefully position two observers on the ground to observe the coordinates x_1 and x_2 of the back and front of the train at some convenient time t . (These two measurements must, of course, be made at the same time t .) In terms of these coordinates, the length l as measured in S is (Fig. 1.10)

$$l = x_2 - x_1.$$



Section 1.12 • Applications of the Lorentz Transformation **29****FIGURE 1.10**

If the two observers measure x_1 and x_2 at the same time ($t_1 = t_2$), then $l = x_2 - x_1$.

Now, consider the following two events, with their coordinates as measured in S .

Event	Description	Coordinates in S
1	Back of train passes first observer	x_1, t_1
2	Front of train passes second observer	$x_2, t_2 = t_1$

We can use the Lorentz transformation to calculate the coordinates of each event as observed in S' .

Event	Coordinates in S'
1	$x'_1 = \gamma(x_1 - vt_1)$
2	$x'_2 = \gamma(x_2 - vt_2)$

(We have not listed the times t'_1 and t'_2 since they don't concern us here.) The difference of these coordinates is

$$x'_2 - x'_1 = \gamma(x_2 - x_1) \quad (1.40)$$

(Notice how the times t_1 and t_2 cancel out because they are equal.) Since the two differences in (1.40) are respectively $l' = l_0$ and l , we conclude that $l_0 = \gamma l$ or

$$l = \frac{l_0}{\gamma}$$

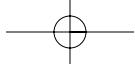
as required.

Example 1.5

Use the Lorentz transformation to rederive the time-dilation formula (1.18).

In our discussion of time dilation we considered two events, a flash and a beep, that occurred at the same place in frame S' ,

$$x'_{\text{flash}} = x'_{\text{beep}}$$



30 Chapter I • The Space and Time of Relativity

The proper time between the two events was the time as measured in S' ,

$$\Delta t_0 = \Delta t' = t'_{\text{beep}} - t'_{\text{flash}}$$

To relate this to the time

$$\Delta t = t_{\text{beep}} - t_{\text{flash}}$$

as measured in S , it is convenient to use the inverse Lorentz transformation (1.38), which gives

$$t_{\text{beep}} = \gamma \left(t'_{\text{beep}} - \frac{vx'_{\text{beep}}}{c^2} \right)$$

and

$$t_{\text{flash}} = \gamma \left(t'_{\text{flash}} - \frac{vx'_{\text{flash}}}{c^2} \right)$$

If we take the difference of these two equations, the coordinates x'_{beep} and x'_{flash} drop out (since they are equal) and we get the desired result,

$$\Delta t = t_{\text{beep}} - t_{\text{flash}} = \gamma(t'_{\text{beep}} - t'_{\text{flash}}) = \gamma \Delta t_0$$

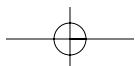
Example 1.6

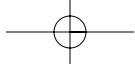
A relativistic snake of proper length 100 cm is moving at speed $v = 0.6c$ to the right across a table. A mischievous boy, wishing to tease the snake, holds two hatchets 100 cm apart and plans to bounce them simultaneously on the table so that the left hatchet lands immediately behind the snake's tail. The boy argues as follows: "The snake is moving with $\beta = 0.6$. Therefore, its length is contracted by a factor

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} = \frac{1}{\sqrt{1 - 0.36}} = \frac{5}{4}$$

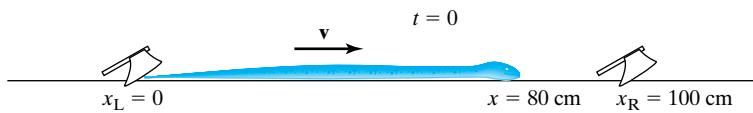
and its length (as measured in my rest frame) is 80 cm. This implies that the right hatchet will fall 20 cm in front of the snake, and the snake will be unharmed." (The boy's view of the experiment is shown in Fig. 1.11.) On the other hand, the snake argues thus: "The hatchets are approaching me with $\beta = 0.6$, and the distance between them is contracted to 80 cm. Since I am 100 cm long, I will be cut in pieces when they fall." Use the Lorentz transformation to resolve this apparent paradox.

Let us choose two coordinate frames as follows: The snake is at rest in frame S' with its tail at the origin $x' = 0$ and its head at $x' = 100$ cm. The two hatchets are at rest in frame S , the left one at the origin $x = 0$ and the right one at $x = 100$ cm.





Section 1.12 • Applications of the Lorentz Transformation 31



As observed in frame S , the two hatchets bounce simultaneously at $t = 0$. At this time the snake's tail is at $x = 0$ and his head must therefore be at $x = 80\text{ cm}$. [You can check this using the transformation $x' = \gamma(x - vt)$; with $x = 80\text{ cm}$ and $t = 0$, you will find that $x' = 100\text{ cm}$, as required.] Thus, as observed in S , the experiment is as shown in Fig. 1.11. In particular, the boy's prediction is correct and the snake is unharmed. Therefore, the snake's argument must be wrong.

To understand what is wrong with the snake's argument, we must examine the coordinates, especially the times, at which the two hatchets bounce, as observed in the frame S' . The left hatchet falls at $t'_L = 0$ and $x'_L = 0$. According to the Lorentz transformation (1.37), the coordinates of this event, as seen in S' , are

$$t'_L = \gamma \left(t_L - \frac{vx_L}{c^2} \right) = 0$$

and

$$x'_L = \gamma(x_L - vt_L) = 0.$$

As expected, the left hatchet falls immediately beside the snake's tail, at time $t'_L = 0$, as shown in Fig. 1.12(a).

On the other hand, the right hatchet falls at $t'_R = 0$ and $x'_R = 100\text{ cm}$. Thus, as seen in S' , it falls at a time given by the Lorentz transformation as

$$t'_R = \gamma \left(t_R - \frac{vx_R}{c^2} \right) = \frac{5}{4} \left(0 - \frac{(0.6c) \times (100\text{ cm})}{c^2} \right) = -2.5\text{ ns}$$

We see that, as measured in S' , the two hatchets *do not fall simultaneously*. Since the right hatchet falls before the left one, it does not necessarily have to hit the snake, even though they were only 80 cm apart (in this frame). In fact, the position at which the right hatchet falls is given by the Lorentz transformation as

$$x'_R = \gamma(x_R - vt_R) = \frac{5}{4}(100\text{ cm} - 0) = 125\text{ cm}$$

and, indeed, the hatchet misses the snake, as shown in Fig. 1.12(b).

The resolution of this paradox and many similar paradoxes is seen to be that two events which are simultaneous as observed in one frame are not necessarily simultaneous when observed in a different frame. As soon as one recognizes that the two hatchets fall at different times in the snake's rest frame, there is no longer any difficulty understanding how they can both miss the snake.

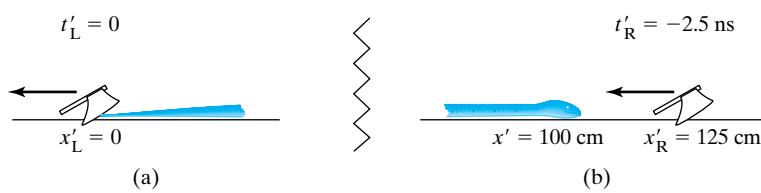
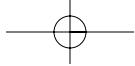


FIGURE 1.11

As seen in the boy's frame S , the two hatchets bounce simultaneously (at $t = 0$) 100 cm apart. Since the snake is 80 cm long, it escapes injury.

FIGURE 1.12

As observed in S' , both hatchets are moving to the left. The right hatchet falls before the left one, and even though the hatchets are only 80 cm apart, this lets them fall at positions that are 125 cm apart.



32 Chapter I • The Space and Time of Relativity

1.13 The Velocity-Addition Formula

In Section 1.3 we discussed the classical velocity-addition formula. This relates the velocity \mathbf{u} of a body or signal, relative to a frame S , and its value \mathbf{u}' relative to a second frame S' , $\mathbf{u} = \mathbf{u}' + \mathbf{v}$, or equivalently,

$$\mathbf{u}' = \mathbf{u} - \mathbf{v} \quad (1.41)$$

Here \mathbf{v} is the velocity of S' relative to S , and the formula asserts that in classical physics, relative velocities add and subtract like vectors. Notice that here, as elsewhere, we use \mathbf{u} and \mathbf{u}' for the velocities of a body or signal relative to the two frames, while \mathbf{v} denotes the relative velocity of the frames themselves.

The Michelson–Morley experiment shows that the classical formula (1.41) cannot be correct, since it would contradict the universality of the speed of light. In this section we use the Lorentz transformation to derive the correct relativistic velocity-addition formula.

Let us imagine some moving object whose velocity we wish to discuss. (For example, this object could be a space rocket, a subatomic particle, or a pulse of light.) We consider two neighboring points on its path, as in Fig. 1.13. We denote by $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$ the coordinates of these two points, as measured in S , and by t_1 and t_2 the times at which the object passes them. The velocity $\mathbf{u} = (u_x, u_y, u_z)$, as measured in S , is then given by

$$u_x = \frac{\Delta x}{\Delta t}, \quad u_y = \frac{\Delta y}{\Delta t}, \quad u_z = \frac{\Delta z}{\Delta t} \quad (1.42)$$

where $\Delta x = x_2 - x_1$, and so on (and these equations may, strictly speaking, be valid only in the limit that the two points are close together, $\Delta t \rightarrow 0$). The velocity \mathbf{u}' relative to S' is defined in the same way, using the coordinates and times measured in S' . That is, $u'_x = \Delta x'/\Delta t'$, and so on.

We can now use the Lorentz transformation to relate the coordinates and times of S to those of S' , and then, using definition (1.42), relate the corresponding velocities. First, according to the Lorentz transformation (1.37)

$$x'_2 = \gamma(x_2 - vt_2), \quad y'_2 = y_2, \quad z'_2 = z_2, \quad t'_2 = \gamma\left(t_2 - \frac{vx_2}{c^2}\right)$$

and

$$x'_1 = \gamma(x_1 - vt_1), \quad y'_1 = y_1, \quad z'_1 = z_1, \quad t'_1 = \gamma\left(t_1 - \frac{vx_1}{c^2}\right)$$

Subtracting these equations, we find that

$$\begin{aligned} \Delta x' &= \gamma(\Delta x - v \Delta t), & \Delta y' &= \Delta y, & \Delta z' &= \Delta z, \\ \Delta t' &= \gamma\left(\Delta t - \frac{v \Delta x}{c^2}\right) \end{aligned}$$

From these we can calculate the components of \mathbf{u}' . First

$$u'_x = \frac{\Delta x'}{\Delta t'} = \frac{\gamma(\Delta x - v \Delta t)}{\gamma(\Delta t - v \Delta x/c^2)}$$

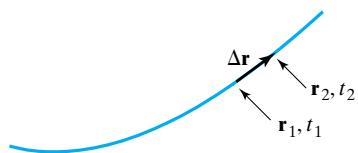
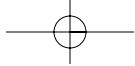


FIGURE 1.13

The velocity of an object is $\mathbf{u} = \Delta \mathbf{r}/\Delta t$.



Section I.13 • The Velocity-Addition Formula 33

or, canceling the factors of γ and dividing top and bottom by Δt ,

$$u'_x = \frac{u_x - v}{1 - u_x v/c^2} \quad (1.43)$$

Similarly,

$$u'_y = \frac{\Delta y'}{\Delta t'} = \frac{\Delta y}{\gamma(\Delta t - v \Delta x/c^2)}$$

or, dividing the top and bottom by Δt ,

$$u'_y = \frac{u_y}{\gamma(1 - u_x v/c^2)} \quad \text{and, similarly,} \quad u'_z = \frac{u_z}{\gamma(1 - u_x v/c^2)} \quad (1.44)$$

Notice that u'_y is not equal to u_y , even though $\Delta y' = \Delta y$; this is because the times $\Delta t'$ and Δt are unequal.

Equations (1.43) and (1.44) are the **relativistic velocity-addition formulas**, or velocity transformation. Notice that if both u and v are much less than c , we can ignore the term $u_x v/c^2$ in the denominators and put $\gamma \approx 1$ to give

$$u'_x \approx u_x - v$$

and

$$u'_y \approx u_y \quad \text{and} \quad u'_z \approx u_z$$

These are, of course, the components of the classical addition formula $\mathbf{u}' = \mathbf{u} - \mathbf{v}$.

The inverse velocity transformation, giving \mathbf{u} in terms of \mathbf{u}' , can be obtained from (1.43) and (1.44) by exchanging primed and unprimed variables and replacing v by $-v$, in the familiar way.

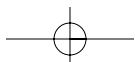
Example 1.7

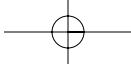
A rocket traveling at speed $0.8c$ relative to the earth shoots forward a beam of particles with speed $0.9c$ relative to the rocket. What is the particles' speed relative to the earth?

Let S be the rest frame of the earth and S' that of the rocket, with x and x' axes both aligned along the rocket's velocity. The relative speed of the two frames is $v = 0.8c$. We are given that the particles are traveling along the x' axis with speed $u' = 0.9c$ (relative to S') and we want to find their speed u relative to S . The classical answer is, of course, that $u = u' + v = 1.7c$, that is, because the two velocities are collinear, u' and v simply add in classical physics.

The correct relativistic answer is given by the inverse of (1.43) (from which we omit the subscripts x , since all velocities are along the x axis).

$$\begin{aligned} u &= \frac{u' + v}{1 + u' v/c^2} \\ &= \frac{0.9c + 0.8c}{1 + (0.9 \times 0.8)} = \frac{1.7}{1.72}c \approx 0.99c \end{aligned} \quad (1.45)$$





34 Chapter I • The Space and Time of Relativity

The striking feature of this answer is that when we “add” $u' = 0.9c$ to $v = 0.8c$ relativistically, we get an answer that is less than c . In fact, it is fairly easy to show that for any value of u' that is less than c , the speed u is also less than c (see Problem 1.47); that is, a particle whose speed is less than c in one frame has speed less than c in any other frame.

Example 1.8

The rocket of Example 1.7 shoots forward a signal (for example, a pulse of light) with speed c relative to the rocket. What is the signal’s speed relative to the earth?

In this case $u' = c$. Thus according to (1.45)

$$u = \frac{u' + v}{1 + u'v/c^2} = \frac{c + v}{1 + v/c} = c \quad (1.46)$$

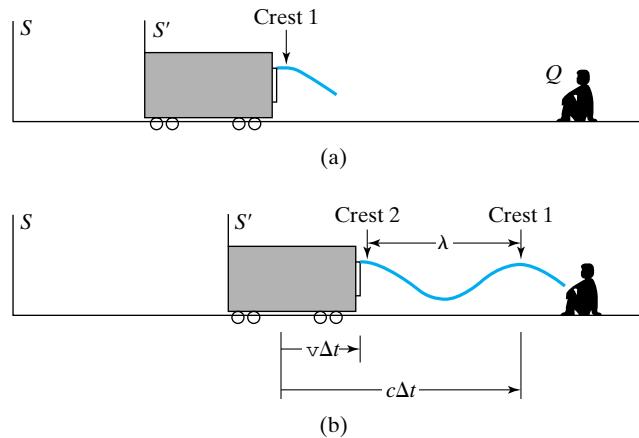
That is, anything that travels at the speed of light in one frame does the same as observed from any other frame. (We have proved this here only for the case that u is in the same direction as v . However, the result is true for any direction; for another example, see Problem 1.48.) We can paraphrase this to say that the speed of light is invariant as we pass from one inertial frame to another. This is, of course, just the second postulate of relativity, which led us to the Lorentz transformation in the first place.

1.14 The Doppler Effect*

*The Doppler effect is an important phenomenon, with applications that range from atomic physics, through police radar traps, to the expanding universe. Nevertheless, we will not be using it again, so, if you are pressed for time, you could omit this section without loss of continuity.

It is well known in classical physics that the frequency (and hence pitch) of sound changes if either the source or receiver is put into motion, a phenomenon known as the **Doppler effect**, after the Austrian physicist Christian Doppler, 1803–1853, who investigated the effect. There is a similar Doppler effect for light: The frequency (and hence color) of light is changed if the source and receiver are put into relative motion. However, the Doppler effect for light differs from that for sound in two important ways. First, since light travels at speed c , one should treat the phenomenon relativistically, and the relativistic Doppler formula differs from its nonrelativistic form because of time dilation. Second, you will probably recall that the nonrelativistic Doppler formula for sound has different forms according to whether the source or receiver is moving. This difference is perfectly reasonable for sound that propagates in air, so that the Doppler shift can legitimately depend on whether it is the source or receiver that moves through the air. On the other hand, we know that light (in the vacuum) has no medium in which it propagates; and, in fact, relativity has shown us that it can make no difference whether it is the source or receiver that is “really” moving. Thus the Doppler formula for light must be the same for a moving source as for a moving receiver. In this respect, the Doppler formula for light is simpler than its nonrelativistic counterpart for sound.

Section 1.14 • The Doppler Effect 35

**FIGURE 1.14**

As seen in frame S , the source moves with speed v and the receiver Q is at rest.

The derivation of the relativistic Doppler formula is very similar to the nonrelativistic argument. Let us consider a train (frame S') that is traveling at speed v relative to the ground (frame S) and whose headlamp is a source of light with frequency f_{src} , as measured in the rest frame of the source. We now imagine an observer Q at rest on the ground in front of the train and wish to find the frequency f_{obs} of the light that Q observes.

We first consider the experiment as seen in frame S , in which it is the source that is moving. In Fig. 1.14(a) and (b) we have shown two successive wave crests (numbered 1 and 2) as they leave the train's headlight. If the time (as measured in S) between the emission of these crests is Δt , then during this time the first crest will move a distance $c \Delta t$ to the right. During this same time the train will advance a distance $v \Delta t$, and the distance λ between successive crests is therefore

$$\lambda = c \Delta t - v \Delta t. \quad (1.47)$$

As seen from S , the distance between successive crests is shortened by $v \Delta t$ as a result of the train's motion. Since the crests are all approaching with speed c and are a distance λ apart, the frequency f_{obs} with which the observer Q receives them is

$$f_{\text{obs}} = \frac{c}{\lambda} = \frac{c}{(c - v)\Delta t} = \frac{1}{(1 - \beta)\Delta t} \quad (1.48)$$

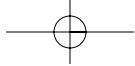
where, as usual, $\beta = v/c$.

Now, Δt is the time between emission of successive crests *as measured in S* . The corresponding time $\Delta t'$ as measured in S' is the proper time between the two events, since they occur at the same place in S' . Therefore,

$$\Delta t = \gamma \Delta t' \quad (1.49)$$

and, from (1.48),

$$f_{\text{obs}} = \frac{1}{(1 - \beta)\gamma \Delta t'} \quad (1.50)$$



36 Chapter I • The Space and Time of Relativity

Finally, the frequency f_{sce} measured at the source is just $f_{\text{sce}} = 1/\Delta t'$. Thus (1.50) implies that

$$f_{\text{obs}} = \frac{f_{\text{sce}}}{(1 - \beta)\gamma} \quad (\text{approaching}) \quad (1.51)$$

where we have added the parenthesis “(approaching)” to emphasize that this formula applies when the source is approaching the observer.

The relativistic formula (1.51) differs from its nonrelativistic counterpart only by the factor of $1/\gamma$, which arose from the time dilation (1.49). In particular, for slow speeds, with $\gamma \approx 1$, we can use the nonrelativistic formula $f_{\text{obs}} = f_{\text{sce}}/(1 - \beta)$, as one might expect.

It is often convenient to rewrite (1.51), replacing the factor of $1/\gamma$ by

$$\frac{1}{\gamma} = \sqrt{1 - \beta^2} = \sqrt{(1 - \beta)(1 + \beta)} \quad (1.52)$$

to give

$$f_{\text{obs}} = \sqrt{\frac{1 + \beta}{1 - \beta}} f_{\text{sce}} \quad (\text{approaching}) \quad (1.53)$$

We will see shortly that this formula, which we have so far derived only for a moving source, in fact holds whether it is the source or observer that is moving.

The formula (1.53) applies to a source that is approaching the observer. If the source is moving away from the observer, we have only to change the sign of v to give

$$f_{\text{obs}} = \sqrt{\frac{1 - \beta}{1 + \beta}} f_{\text{sce}} \quad (\text{receding}) \quad (1.54)$$

The formulas (1.53) and (1.54) are easy to memorize. In particular, it is easy to remember which is which, since both numerator and denominator lead to the expected *rise* in frequency when source and observer are approaching each other, and the expected *drop* when they are receding.

We have here analyzed only the cases that the source moves directly toward or away from the observer. The case that the source moves obliquely to the observer is more complicated and is discussed in Problem 1.53.

An important example of the Doppler effect is the famous “redshift” of the light from distant stars. A star emits and absorbs light at certain frequencies that are characteristic of the elements in the star. Thus, by analyzing the spectrum of light from a star, one can identify which elements it contains. Once these elements are identified, one can go further. By seeing whether the characteristic frequencies are shifted up or down (as compared to those from a source at rest in the observatory) one can tell whether the star is moving toward or away from us. The American astronomer Edwin Hubble found that the light from distant galaxies is shifted down in frequency, or redshifted (since red is at the low frequency end of the visible spectrum), indicating that most galaxies are moving away from us. Hubble also found that the speeds of recession of galaxies are roughly proportional to their distances from us, a discovery now called Hubble’s law. This implied that the universe is expanding uniform-

ly; it also provided a convenient way to find the distance of many galaxies, since measurement of a Doppler shift is usually much easier than the direct measurement of a distance.

Example 1.9

It is found that light from a distant galaxy is shifted down in frequency (red-shifted) by a factor of 3; that is, $f_{\text{obs}}/f_{\text{sce}} = 1/3$. Is the galaxy approaching us or receding? And what is its speed?

Since the observed frequency f_{obs} is less than f_{sce} , the galaxy is receding, and we use (1.54) to give

$$\sqrt{\frac{1 - \beta}{1 + \beta}} = \frac{1}{3}$$

or, solving for β ,

$$\beta = 0.8$$

That is, the galaxy is receding from us at $0.8c$.

The Doppler effect for light has many uses here on the earth. One application that is all too familiar to many of us is the Doppler radar used by police to measure our car's speed. This actually involves two Doppler shifts: As "seen" by our car, the radar is shifted up in frequency as we approach the gun, and the tiny electric currents induced in the car's body are of slightly higher frequency than that of the gun. This means that the reflected signal sent back by our car has the same higher frequency, and this signal (from our moving car) is then raised again as seen by the radar receiver. The total shift is very small (since $\beta \ll 1$), but is easily measured and immediately converted to give the car's speed. In atomic physics a process called **laser cooling** exploits the Doppler effect to slow the thermal motions of the atoms or molecules in a gas. The same Doppler effect is also a significant nuisance to atomic spectroscopists, since it broadens the spectral lines that they wish to measure, as the following example illustrates.

Example 1.10

The atoms in hot sodium vapor give out light of wavelength $\lambda_{\text{sce}} = 589 \text{ nm}$ (measured in the atoms' rest frame). Since atoms in a vapor move randomly with speeds up to 300 m/s and even higher, this light is observed with various different Doppler shifts, depending on the atoms' speeds and directions. Taking 300 m/s as the atoms' maximum speed, find the range of wavelengths observed.

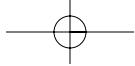
The minimum and maximum frequencies observed come from atoms moving directly away from and toward the observer with speed $v = 300 \text{ m/s}$ or $\beta = 10^{-6}$. Since β is so small, we can ignore the factor of γ in (1.51) and the extreme frequencies are given by

$$f_{\text{obs}} = \frac{f_{\text{sce}}}{1 \pm \beta}$$

Edwin Hubble (1889–1953, American)



Raised in midwest America, the young Hubble wanted to study astronomy, but, at the insistence of his father, obtained a law degree at Oxford University. After his father died, Hubble returned to college to study astronomy and quickly became one of the world's foremost observational astronomers. Working with what was then the largest telescope in the world, the 100-inch-diameter reflecting telescope at Mt. Wilson Observatory in California, Hubble established the two most important results of twentieth-century astronomy. He demonstrated that our Milky Way galaxy is only one of myriads of galaxies in the universe. And he showed that distant galaxies are receding from us with a speed proportional to their distance, thus establishing that the universe is expanding.



38 Chapter I • The Space and Time of Relativity

Since $\lambda = c/f$ (both for source and observer), this implies maximum and minimum wavelengths given by

$$\lambda_{\text{obs}} = \lambda_{\text{sce}}(1 \pm \beta)$$

We can write this as

$$\lambda_{\text{obs}} = \lambda_{\text{sce}} \pm \Delta\lambda$$

where the maximum shift $\Delta\lambda$ in the wavelength is

$$\Delta\lambda = \beta\lambda_{\text{sce}} = (10^{-6}) \times (589 \text{ nm}) \approx 6 \times 10^{-4} \text{ nm}$$

This is a very small shift of wavelength as we should have expected, since v is so small compared to c . Nevertheless, such a shift is easily observed with a good spectrometer. It means that what would otherwise be observed as a sharp spectral line with wavelength λ_{sce} is smeared out between $\lambda_{\text{sce}} \pm \Delta\lambda$. This phenomenon is called **Doppler broadening** and is one of the problems that has to be overcome in precise measurement of wavelengths.

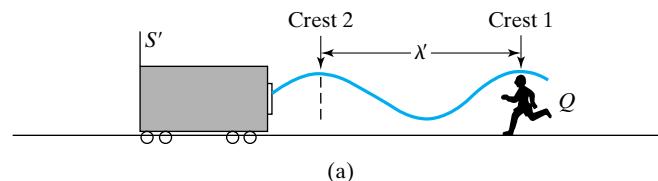
We mentioned earlier that the relativistic Doppler shift for light must be the same whether we view the source as moving and the observer at rest, or vice versa. We check this in our final example.

Example 1.11

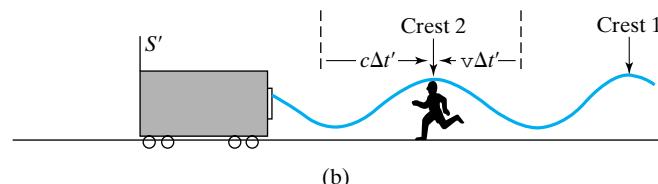
Rederive the Doppler formula (1.53) working in the rest frame S' of the source (that is, taking the view that the observer is moving).

We consider again two successive wave crests, but examine their *reception* by the observer Q , as shown in Fig. 1.15. As measured in S' , the distance between the two crests is the wavelength $\lambda' = c/f_{\text{sce}}$ (since f_{sce} is the frequency measured in S'), and the time between Q 's meeting the crests we denote by $\Delta t'$. During the time $\Delta t'$, crest 2 moves a distance $c \Delta t'$ to the right and the observer Q moves a distance $v \Delta t'$ to the left. The sum of these two distances is just λ' ,

$$c \Delta t' + v \Delta t' = \lambda' = \frac{c}{f_{\text{sce}}}$$



(a)



(b)

FIGURE 1.15

As seen in S' , the source is stationary with frequency f_{source} , and the observer Q is moving at speed v to the left.

from which we find that

$$\Delta t' = \frac{c}{(c + v)f_{\text{sce}}} = \frac{1}{(1 + \beta)f_{\text{sce}}} \quad (1.55)$$

Now, the frequency with which Q observes wave crests is

$$f_{\text{obs}} = \frac{1}{\Delta t} \quad (1.56)$$

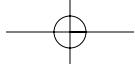
where Δt is the time between arrival of the two successive crests *as measured by the observer Q* . Since these two events occur at the same place in the observer's frame S , Δt is the proper time and $\Delta t' = \gamma \Delta t$. Substituting into (1.56) and then using (1.55), we find that

$$f_{\text{obs}} = \frac{\gamma}{\Delta t'} = \gamma(1 + \beta)f_{\text{sce}} \quad (1.57)$$

Apart from the factor γ , this is the nonrelativistic formula for a moving receiver. For our present purposes, the important point is that we can replace γ using (1.52), and after a little algebra, we obtain exactly our previous answer (1.53), as you should check for yourself. As anticipated, the relativistic Doppler shift for light is the same for a moving observer as for a moving source, and depends only on their relative velocity v .

CHECKLIST FOR CHAPTER I

CONCEPT	DETAILS
Relativity of measurements	Sec. 1.1
The classical velocity-addition formula	$u = u' + v$ (1.1)
Invariance of Newton's laws in classical physics	Sec. 1.3
Noninvariance of the speed of light in classical physics	Sec. 1.4
Michelson–Morley experiment*	(Sec. 1.5)
The postulates of relativity	Definition of inertial frames The two postulates (Sec. 1.6)
Time dilation	$\Delta t = \gamma \Delta t_0$ (1.18)
The speed limit for the relative motion of inertial frames is c	Sec. 1.8
Proper time	Sec. 1.8
Lorentz–Fitzgerald contraction	$l = l_0/\gamma$ (1.29)
Proper length	Sec. 1.10
The Galilean transformation	$x' = x - vt, y' = y, z' = z, t' = t$ (1.31)
The Lorentz transformation	$x' = \gamma(x - vt), y' = y, z' = z, t' = \gamma(t - vx/c^2)$ (1.37)
Relativistic velocity addition	$u'_x = \frac{u_x - v}{1 - u_x v/c^2}, u'_y = \frac{u_y}{\gamma(1 - u_x v/c^2)}, u'_z = \frac{u_z}{\gamma(1 - u_x v/c^2)}$ (1.43) & (1.44)
The Doppler effect*	$f_{\text{obs}} = f_{\text{sce}} \sqrt{(1 + \beta)/(1 - \beta)}$ (1.53)



40 Chapter I • The Space and Time of Relativity

PROBLEMS FOR CHAPTER I

The problems for each chapter are arranged according to section number. A problem listed for a given section requires an understanding of that section and earlier sections, but not of later sections. Within each section, problems are listed in approximate order of difficulty. A single dot (•) indicates straightforward problems involving just one main concept and sometimes requiring no more than substitution of numbers in the appropriate formula. Two dots (••) identify problems that are slightly more challenging and usually involve more than one concept. Three dots (•••) indicate problems that are distinctly more challenging, either because they are intrinsically difficult or involve lengthy calculations. Needless to say, these distinctions are hard to draw and are only approximate.

Answers to odd-numbered problems are given at the back of the book.

SECTION 1.2 (The Relativity of Orientation and Origin)

- 1.1** •• At time $t = 0$, a block is released from the point O on the slope shown in Fig. 1.16. The block accelerates down the slope, overcoming the sliding friction (coefficient μ). **(a)** Choose axes Oxy as shown, and resolve the equation $\Sigma F = ma$ into its x and y components. Hence find the block's position (x, y) as a function of time, and the time it takes to reach the bottom. **(b)** Carry out the solution using axes $Ox'y'$, with Ox' horizontal and Oy' vertical, and show that you get the same final answer. Explain why the solution using these axes is less convenient.

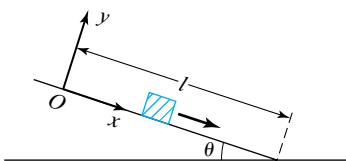


FIGURE 1.16
(Problem 1.1)

- 1.2** •• A block slides down the slope of Fig. 1.17 from O with initial speed v_0 . The sliding friction (coefficient μ) brings the block to rest in a time T . **(a)** Using the axes shown, find T . **(b)** Solve this problem using axes $Ox'y'$ with Ox' horizontal and Oy' vertical, and explain why the solution in this frame is less convenient (although it produces the same final answer, of course).

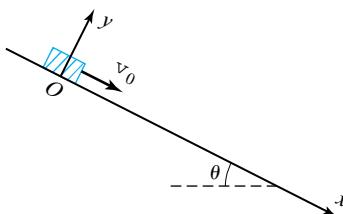


FIGURE 1.17
(Problem 1.2)

- 1.3** ••• At time $t = 0$ a ball is thrown with speed v_0 at an angle θ above a slope that is itself inclined at an angle

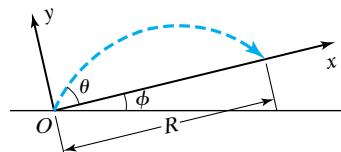


FIGURE 1.18
(Problem 1.3)

ϕ above the horizontal (Fig. 1.18). **(a)** Choosing axes as shown, write down the components of the ball's initial velocity v_0 and its acceleration \mathbf{g} . Hence find the ball's position (x, y) as a function of time and show that its range up the slope is

$$R = \frac{2v_0^2 \sin \theta \cos(\theta + \phi)}{g \cos^2 \phi}$$

Note that when the ball lands, $y = 0$. **(b)** Show that you get the same final answer if you use axes $Ox'y'$ with Ox' horizontal and Oy' vertical. Discuss briefly the merits of the two choices of axes. (You can find several useful trig identities in Appendix B.)

SECTION 1.3 (Moving Reference Frames)

- 1.4** • A physics lecture demonstration uses a small cannon mounted on a cart that moves at constant velocity \mathbf{v} across the floor. At what angle θ should the cannon point (measured from the horizontal floor of the cart) if the cannonball is to land back in the mouth of the cannon? Explain clearly your choice of frame of reference.

- 1.5** • Two students are riding on a flatcar, traveling at speed u to the right. One is standing at the right end of the car, holding a small vertical hoop at a height h above the car's floor. The other has a catapult with which he plans to shoot a pellet from the car's floor and through the hoop. He is aiming the catapult at an angle θ above the floor of the car. If he wants the pellet to be traveling horizontally when it goes through the hoop, with what speed (relative to the car) should he fire the pellet? How far horizontally should he be from his companion? Explain your choice of reference frame.

- 1.6** • Consider a head-on, elastic collision between two bodies whose masses are m and M , with $m \ll M$. It is well known that if m has speed v_0 and M is initially at rest, m will bounce straight back with its speed unchanged, while M will remain at rest (to an excellent approximation). Use this fact to predict the final velocities if M approaches with speed v_0 and m is initially at rest.

[HINT: Consider the reference frame attached to M .]

- 1.7** • Use the method of Problem 1.6 to predict the final velocities if two bodies of masses m and M , with $m \ll M$, approach one another both traveling at speed v_0 (relative to the lab) and undergo a head on, elastic collision.

- 1.8** •• A policeman is chasing a robber. Both are in cars traveling at speed v and the distance between them

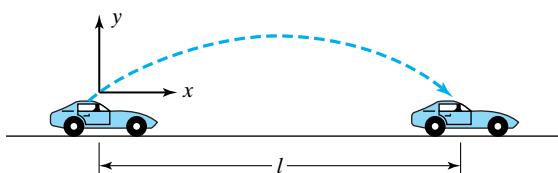
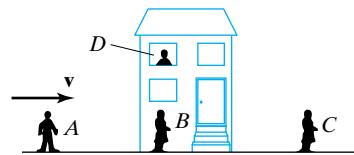


FIGURE 1.19
(Problem 1.8)

is l . The policeman wishes to shoot the robber with a gun whose muzzle velocity is u_0 . At what angle θ above the horizontal should he point his gun? First solve this problem using coordinates traveling with the policeman, as shown in Fig. 1.19. Then sketch the solution using coordinates fixed to the ground; is the angle of the gun the same as the angle of the bullet's initial velocity in this frame? (The advantages of the first frame are not overwhelming; nevertheless, it is clearly the natural choice for the problem.)

SECTION 1.4 (Classical Relativity and the Speed of Light)

- 1.9** • Let us assume the classical ideas of space and time are correct, so that there could only be one frame, the “ether frame,” in which light traveled at the same speed c in all directions. It seemed unlikely that the earth would be exactly at rest in this frame and one might reasonably have guessed that the earth’s speed v relative to the ether frame would be at least of the order of our orbital speed around the sun ($v \approx 3 \times 10^4$ m/s). (a) What would be the observed speed (on earth) of a light wave traveling parallel to \mathbf{v} ? (Give your answer in terms of c and v , and then substitute numerical values.) (b) What if it were traveling antiparallel to \mathbf{v} ? (c) What if it were traveling perpendicular to \mathbf{v} (as measured on earth)? The accepted value of c is 2.9979×10^8 m/s (to five significant figures).
- 1.10** •• At standard temperature and pressure sound travels at speed $u = 330$ m/s relative to the air through which it propagates. Four students, A , B , C , D , position themselves as shown in Fig. 1.20, with A , B , C in a straight line and D vertically above B . A steady wind is blowing with speed $v = 30$ m/s along the line ABC . If B fires a revolver, what are the speeds with which the sound will travel to A , C , and D (in the reference frame of the observers)? Discuss whether the differences in your answers could be detected.



- 1.11** ••• It is well known that the speed of sound in air is $u = 330$ m/s at standard temperature and pressure. What this means is that sound travels at speed u in all directions in the frame S where the air is at rest. In any

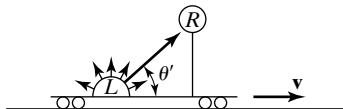


FIGURE 1.21
(Problem 1.11)

other frame S' , moving relative to S , its speed is *not* u in all directions. To verify this, some students set up a loudspeaker L and receiver R on an open flatcar, as in Fig. 1.21; by connecting the electrical signals from L and R to an oscilloscope, they can measure the time for a sound to travel from L to R and hence find its speed u' (relative to the car). (a) Derive an expression for u' in terms of u , v , and θ' , where v is the car’s speed through the air and θ' is the angle between \mathbf{v} and \mathbf{LR} . (We call this θ' since it is the angle between \mathbf{v} and \mathbf{u}' , the velocity of the sound measured in the frame of the car.)

[*HINT:* Draw a velocity-addition triangle to represent the relation $\mathbf{u} = \mathbf{u}' + \mathbf{v}$. The law of cosines should give you a quadratic equation for u' .]

(b) If the students vary the angle θ' from 0 to 180° , what are the largest and smallest values of u' ? (c) If v is about 15 m/s (roughly 30 mi/h), what will be the approximate percent variation in u' ? Would this be detectable?

SECTION 1.5 (The Michelson–Morley Experiment*)

- 1.12** • In the discussion of the Michelson–Morley experiment, we twice used the binomial approximation

$$(1 - x)^n \approx 1 - nx \quad (1.58)$$

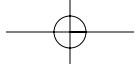
which holds for any number n and any x much smaller than 1 (that is, $|x| \ll 1$). (In the examples, n was -1 and $-\frac{1}{2}$, and $x = \beta^2$ was of order 10^{-8} .) The binomial approximation is frequently useful in relativity, where one often encounters expressions of the form $(1 - x)^n$ with x small. Make a table showing $(1 - x)^n$ and its approximation $1 - nx$ for $n = -\frac{1}{2}$ and $x = 0.5, 0.1, 0.01$, and 0.001 . In each case find the percentage by which the approximation differs from the exact result.

- 1.13** • Do the same tasks as in Problem 1.12, but for the case $n = 2$. In this case give an exact expression for the difference between the exact and approximate forms. Explain why the approximation gets better and better as $x \rightarrow 0$.

- 1.14** • Use the binomial approximation (1.58) (Problem 1.12) to evaluate $(1 - 10^{-20})^{-1} - 1$. Can you evaluate this directly on your calculator?

- 1.15** • Tom Sawyer and Huck Finn can each row a boat at 5 ft/s in still water. Tom challenges Huck to a race in which Tom is to row the 2000 ft across the Mississippi to a point exactly opposite their starting point and back again, while Huck rows to a point 2000 ft directly downstream and back up again. If the Mississippi flows at 3 ft/s, which boy wins and by how long?

- 1.16** • An airline, all of whose planes fly with an airspeed of 200 mi/h, serves three cities, A , B , and C , where B



42 Chapter I • The Space and Time of Relativity

is 320 mi due east of A , and C is the same distance due north of A . On a certain day there is a steady wind of 120 mi/h from the east. (a) What is the time needed for a round trip from A to B and back? (b) What is it from A to C and back?

- 1.17** • In one of the early (1881) versions of Michelson's interferometer, the arms were about 50 cm long. What would be the expected shift ΔN when he rotated the apparatus through 90° , assuming that $\lambda = 590 \text{ nm}$ and that the expected speed of the earth relative to the ether was $3 \times 10^4 \text{ m/s}$? (In this case the expected shift was so small that no one regarded his failure to observe a shift as conclusive.)

- 1.18** •• One of the difficulties with the Michelson–Morley experiment is that several extraneous effects (mechanical vibrations, variations in temperature, etc.) can produce unwanted shifts in the interference pattern, masking the shift of interest. Suppose, for example, that during the experiment the temperature of one arm of the interferometer were to rise by ΔT . This would increase the arm's length by $\Delta l = \alpha l \Delta T$, where α is the arm's coefficient of expansion. Prove that this temperature change would, by itself, cause a shift $\Delta N = 2\alpha l \Delta T / \lambda$. For the dimensions given in Problem 1.17 and taking $\alpha \approx 10^{-5}/^\circ\text{C}$ (the coefficient for steel) and $\Delta T \approx 0.01^\circ\text{C}$, show that the resulting shift is $\Delta N \approx 0.2$, much larger than the expected shift due to the earth's motion. Obviously a successful experiment requires careful temperature control!

SECTIONS 1.8 and 1.9 (The Relativity of Time and Evidence for Time Dilation)

- 1.19** • An athlete runs the 100-meter dash at 10 m/s. How much will her watch gain or lose, as compared to ground-based clocks, during the race?

[HINT: You will need to use the binomial approximation (1.6).]

- 1.20** • A space vehicle travels at 100,000 m/s (about 200,000 mi/h) relative to the earth. How much time will its clocks gain or lose, as compared to earth-based clocks, in a day?

- 1.21** • (a) What must be one's speed, relative to a frame S , in order that one's clocks will lose 1 second per day as observed from S ? (b) What if they are to lose 1 minute per day?

- 1.22** • A space explorer sets off at a steady $v = 0.95c$ to a distant star. After exploring the star for a short time he returns at the same speed and gets home after a total absence of 80 years (as measured by earthbound observers). How long do his clocks say that he was gone, and by how much has he aged? Note: This is the twin paradox discussed in Example 1.3. It is easy to get the right answer by judicious insertion of a factor γ in the right place, but to understand the result you need to recognize that it involves three inertial frames: the earthbound frame S , the frame S' of the outward-bound rocket, and the frame S'' of the returning rocket. You can write down the time-dilation

formula (1.14) for each of the two halves of the journey, and add these to give the desired relation. (Notice that the experiment is not symmetrical between the explorer and his friends who stay behind on earth — the earthbound clocks stay at rest in a single inertial frame, but the rocket's clock and crew occupy at least two different frames. This is what allows the result to be unsymmetrical.)

- 1.23** • (When he returns his Hertz rent-a-rocket after one week's cruising in the galaxy, Mr. Spock is shocked to be billed for three weeks' rental. Assuming that he traveled straight out and then straight back, always at the same speed, how fast was he traveling? (See the note in Problem 1.22.)

- 1.24** •• (a) Use the binomial approximation (1.6) to prove the following useful approximation:

$$\gamma \approx 1 + \frac{1}{2} \beta^2$$

valid when $\beta \ll 1$. (b) Derive a corresponding approximation for $1/\gamma$. (c) When β is close to 1 (v close to c) these approximations are, of course, useless; in this case show that if $\beta = 1 - \varepsilon$, with $\varepsilon \ll 1$, then $\gamma \approx 1/\sqrt{2\varepsilon}$.

- 1.25** •• Two perfectly synchronized clocks A and B are at rest in S , a distance d apart. If we wanted to verify that they really are synchronized, we might try using a third clock, C . We could bring C close to A and check that A and C agree, then move C over to B and check the agreement of B and C . Unfortunately, this procedure is suspect since clock C will run differently while it is being moved. (a) Suppose that A and C are found to be in perfect agreement and that C is then moved at constant speed v to B . Derive an expression for the disagreement τ between B and C , in terms of v and d . What is τ if $v = 300 \text{ m/s}$ and $d = 1000 \text{ km}$? (b) Show that the method can nevertheless be made satisfactory to *any desired* accuracy by moving clock C slowly enough; that is, we can make τ as small as we please by choosing v sufficiently small.

- 1.26** •• A group of π mesons (pions) is observed traveling at speed $0.8c$ in a particle-physics laboratory. (a) What is the factor γ for the pions? (b) If the pions' proper half-life is $1.8 \times 10^{-8} \text{ s}$, what is their half-life as observed in the lab frame? (c) If there were initially 32,000 pions, how many will be left after they have traveled 36 m? (d) What would be the answer to (c) if one ignored time dilation?

- 1.27** •• Muons are subatomic particles that are produced several miles above the earth's surface as a result of collisions of cosmic rays (charged particles, such as protons, that enter the earth's atmosphere from space) with atoms in the atmosphere. These muons rain down more-or-less uniformly on the ground, although some of them decay on the way since the muon is unstable with a proper half-life of about $1.5 \mu\text{s}$. ($1 \mu\text{s} = 10^{-6} \text{ s}$.) In a certain experiment a muon detector is carried in a balloon to an altitude of 2000 m, and in the course of 1 hour it registers 650 muons traveling at $0.99c$ toward the earth. If an identical detector remains at sea level, how many muons

would you expect it to register in 1 hour? (Remember that after n half-lives the number of muons surviving from an initial sample of N_0 is $N_0/2^n$, and don't forget about time dilation). This was essentially the method used in the first tests of time dilation, starting in the 1940's.

- 1.28** ••• Time dilation implies that if a clock moves relative to a frame S , careful measurements made by observers in S [as in Fig. 1.22(a), for example] will find that the clock runs slow. This is not at all the same thing as saying that a single observer in S will *see* the clock running slow; and the latter statement is, in fact, not always true. To understand this, remember that what we *see* is determined by the light as it arrives at our eyes. Consider the observer Q in Fig. 1.22(b) and suppose that as the clock moves from A to B , it registers the passage of a time τ_0 . As measured in S , the time between these two events ("clock at A " and "clock at B ") is of course $\tau = \gamma\tau_0$. However, B is closer to Q than A is; thus light from the clock when at B will reach Q in a shorter time than will light from the clock when at A . Therefore, the time τ_{sec} between Q 's *seeing* the clock at A and *seeing* it at B is less than τ . (a) Prove that in fact

$$\tau_{\text{sec}} = \tau(1 - \beta) = \tau_0 \sqrt{\frac{1 - \beta}{1 + \beta}}$$

(Prove both equalities.) Since τ_{sec} is less than τ_0 , the observer Q actually sees the clock running fast.

(b) What will Q see once the clock has passed her? That is, find the new value of τ_{sec} when the clock is moving away from Q .

Your answers here are closely related to the Doppler effect discussed in Section 1.14. The moral of this problem is that one must be very careful how one states (and thinks about) time dilation. It is safe to say "moving clocks are observed to run slow" [where to "observe" means to "measure carefully" as in Fig. 1.22(a)], but it is certainly wrong to say "moving clocks are *seen* to run slow."

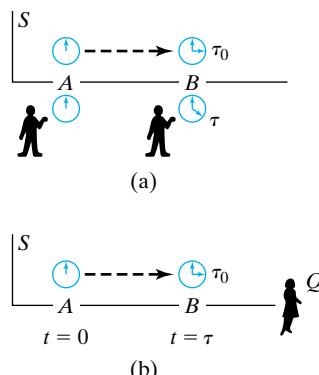


FIGURE 1.22

(Problem 1.28) (a) Two observers at rest in frame S at A and B time the moving clock as it passes them; they find the dilated time $\tau = \gamma\tau_0$. (b) The single observer Q sees the moving clock at A and B by means of light that has traveled different distances, AQ and BQ .

SECTION 1.10 (Length Contraction)

- 1.29** • A rocket of proper length 40 m is observed to be 32 m long as it rushes past the earth. What is its speed relative to the earth?
- 1.30** • A relativistic conveyor belt is moving at speed $0.5c$ relative to frame S . Two observers standing beside the belt, 10 ft apart as measured in S , arrange that each will paint a mark on the belt at exactly the same instant (as measured in S). How far apart will the marks be as measured by observers on the belt?
- 1.31** • A rigid spherical ball (rest frame S) is observed from a frame S' that travels with speed $0.5c$ relative to frame S . Describe the ball's shape as measured by observers in S' .
- 1.32** • Consider the experiment of Problem 1.26 from the point of view of the pions' rest frame. In part (c) how far (as "seen" by the pions) does the laboratory move, and how long does this take? How many pions remain at the end of this time?
- 1.33** • A meter stick is moving with speed $0.8c$ relative to a frame S . (a) What is the stick's length, as measured by observers in S , if the stick is parallel to its velocity \mathbf{v} ? (b) What if the stick is perpendicular to \mathbf{v} ? (c) What if the stick is at 60° to \mathbf{v} , as seen in the stick's rest frame?

[HINT: You can imagine that the meterstick is the hypotenuse of a 30–60–90 triangle of plywood.]

- (d) What if the stick is at 60° to \mathbf{v} , as measured in S ?
- 1.34** •• Like time dilation, the Lorentz contraction cannot be *seen* directly (that is, perceived by the normal process of vision). To understand this claim, consider a rod of proper length l_0 moving relative to S . Careful measurements made by observers in S [as in Fig. 1.23(a), for example] will show that the rod has

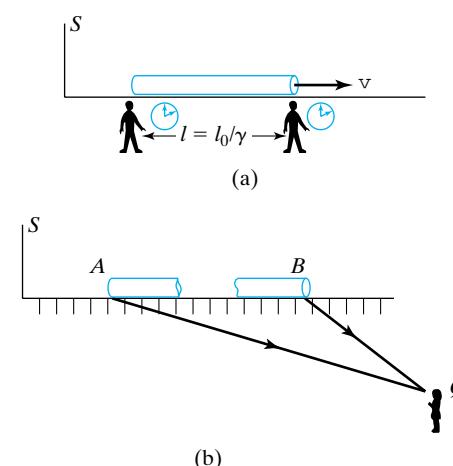


FIGURE 1.23

(Problem 1.34) (a) One can measure the Lorentz-contracted length $l = l_0/\gamma$ using two observers to record the positions of the front and back at the same instant. (b) What a single observer sees is determined by light that left the rod at different times.

44 Chapter I • The Space and Time of Relativity

the contracted length $l = l_0/\gamma$. But now consider what is seen by observer Q in Fig. 1.23(b) (with Q to the right of points A and B). What Q sees at any one instant is determined by the light entering her eyes *at that instant*. Now, consider the light reaching Q at one instant from the front and back of the rod. **(a)** Explain why these two rays must have left the rod (from points A and B) at *different times*. If the x axis has a graduated scale as shown, Q sees (and a photograph would record) a rod extending from A to B ; that is, Q sees a rod of length AB . **(b)** Prove that Q sees a rod that is *longer* than l . (In fact, at certain speeds it is even seen to be longer than l_0 , and the Lorentz contraction is distorted into an expansion.) **(c)** Prove that once it has passed her, Q will see the rod to be shorter than l .

SECTIONS 1.11 and 1.12 (The Lorentz Transformation and Applications)

- 1.35** • The two frames S and S' are in the standard configuration (origins coincident at $t = t' = 0$, x and x' axes parallel, and relative velocity along Ox). Their relative speed is $0.5c$. An event occurs on the x axis at $x = 10$ light-seconds (a light-second is the distance traveled by light in one second, $1 c \cdot \text{sec} = 3 \times 10^8 \text{ m}$) at time $t = 4$ s in the frame S . What are its coordinates x', y', z', t' as measured in S' ?
- 1.36** • The Lorentz transformation (1.37) consists of four equations giving x', y', z', t' in terms of x, y, z, t . Solve these equations to give x, y, z, t in terms of x', y', z', t' . Show that you get the same result by interchanging primed and unprimed variables and replacing v by $-v$.
- 1.37** • Give in detail the derivation of the Lorentz transformation (1.36) for t' , starting from equations (1.34) and (1.35).
- 1.38** • Two inertial frames S and S' are in the standard configuration, with relative velocity \mathbf{v} along the line of the x and x' axes. Consider any two events, 1 and 2. **(a)** From the Lorentz transformation (1.37), derive expressions for the separations $\Delta x', \Delta y', \Delta z', \Delta t'$ (where $\Delta x' = x'_2 - x'_1$, etc.) in terms of $\Delta x, \Delta y, \Delta z, \Delta t$. (Notice how the transformation of $\Delta x, \Delta y, \Delta z, \Delta t$ is identical to that of x, y, z, t .) **(b)** If $\Delta x = 0$ and $\Delta t = 4$ s, whereas $\Delta t' = 5$ s, what is the relative speed v , and what is $\Delta x'$?
- 1.39** •• The frames S and S' are in the standard configuration with relative velocity $0.8c$ along Ox . **(a)** What are the coordinates (x_1, y_1, z_1, t_1) in S of an event that occurs on the x' axis with $x'_1 = 1500 \text{ m}$, $t'_1 = 5 \mu\text{s}$? **(b)** Answer the same for a second event on the x' axis with $x'_2 = -1500 \text{ m}$, $t'_2 = 10 \mu\text{s}$. **(c)** What are the time intervals (Δt and $\Delta t'$) between the two events, as measured in S and S' ?
- 1.40** •• In a frame S , two events have spatial separation $\Delta x = 600 \text{ m}$, $\Delta y = \Delta z = 0$, and temporal separation $\Delta t = 1 \mu\text{s}$. A second frame S' is moving along Ox with nonzero speed v and $O'x'$ parallel to Ox . In S' it is found that the spatial separation $\Delta x'$ is also 600 m . What are v and $\Delta t'$?

- 1.41** •• Observers in a frame S arrange for two simultaneous explosions at time $t = 0$. The first explosion is at the origin ($x_1 = y_1 = z_1 = 0$) while the second is on the positive x axis 4 light years away ($x_2 = 4c \cdot \text{years}$, $y_2 = z_2 = 0$). **(a)** Use the Lorentz transformation to find the coordinates of these two events as observed in a frame S' traveling in the standard configuration at speed $0.6c$ relative to S . **(b)** How far apart are the two events as measured in S' ? **(c)** Are the events simultaneous as observed in S' ?

- 1.42** •• A traveler in a rocket of length $2d$ sets up a coordinate system S' with origin O' anchored at the exact middle of the rocket and the x' axis along the rocket's length. At $t' = 0$ she ignites a flashbulb at O' . **(a)** Write down the coordinates x'_F, t'_F and x'_B, t'_B for the arrival of the light at the front and back of the rocket. **(b)** Now consider the same experiment as observed in a frame S relative to which the rocket is traveling at speed v (with S and S' arranged in the standard configuration). Use the Lorentz transformation to find the coordinates x_F, t_F and x_B, t_B of the arrival of the two signals. Explain clearly why the two times are not equal in frame S , although they were in S' . (This illustrates how two events that are simultaneous in S' are not necessarily simultaneous in S .)

- 1.43** •• Consider the relativistic snake of Example 1.6, but let the numbers be as follows: The snake has speed $0.6c$ and proper length of 100 cm (as before), but the boy holds the two hatchets 80 cm apart. **(a)** Show that with these lengths the experiment can be seen as a test of relativity, since the snake will be unhurt if relativity is right (and the boy times things correctly), whereas the snake will definitely be hurt if the classical ideas of space and time are correct. (Naturally, relativity is correct and the snake is unharmed.) **(b)** Use the Lorentz transformation to find the positions and times of the falling of the two hatchets as measured by the snake, and use these to verify that it is unharmed. (Assume the boy bounces the hatchets at $t = 0$, at which time the snake's tail is at the common origin.)

- 1.44** ••• **(a)** Consider two frames S and S' that differ only by a rotation in which the x and y axes were rotated clockwise through an angle θ to become x' and y' . Prove that

$$x' = x \cos \theta - y \sin \theta \quad \text{and} \quad y' = y \cos \theta + x \sin \theta$$

(and $z' = z$ and $t' = t$). **(b)** Prove that the standard Lorentz transformation can be written as

$$x' = x \cosh \phi - ct \sinh \phi$$

and

$$ct' = ct \cosh \phi - x \sinh \phi$$

(and $y' = y$ and $z' = z$) where $\phi = \tanh^{-1}(v/c)$. Except that the trig functions \cos and \sin are replaced by the hyperbolic functions \cosh and \sinh (and that one term has changed sign), the Lorentz transformation does to x and ct just what a rotation does to x and y . This is our first indication that x, y, z, ct should be re-

Problems for Chapter I **45**

garded as the four coordinates in some kind of four-dimensional space-time.

SECTION 1.13 (The Velocity-Addition Formula)

- 1.45** • A rocket (rest frame S') traveling at speed $v = 0.5c$ relative to the earth (rest frame S) shoots forward bullets traveling at speed $u' = 0.6c$ relative to the rocket. What is the bullets' speed u relative to the earth?

- 1.46** • As seen from earth (rest frame S) two rockets A and B are approaching in opposite directions, each with speed $0.9c$ relative to S . Find the velocity of rocket B as measured by the pilot of rocket A .

[*HINT:* Consider a coordinate system S' traveling with rocket A ; your problem is then to find the velocity \mathbf{u}' of rocket B relative to S' , knowing its velocity \mathbf{u} relative to S .]

- 1.47** •• Using the velocity-addition formula, one can prove the following important theorem: If a body's speed u relative to an inertial frame S is less than c , its speed u' relative to any other inertial frame S' is also less than c . In this problem you will prove this result for the case that all velocities are in the x direction.

Suppose that S' is moving along the x axis of frame S with speed v . Suppose that a body is traveling along the x axis with velocity u relative to S . (We can let u be positive or negative, so that the body can be traveling either way.) (a) Write down the body's velocity u' relative to S' . For a fixed positive v (less than c , of course), sketch a graph of u' as a function of u in the range $-c < u < c$. (b) Hence prove that for any u with $-c < u < c$, it is necessarily true that $-c < u' < c$.

- 1.48** •• Suppose that as seen in a frame S , a signal (a pulse of light, for example) has velocity c along the y axis (that is, $u_x = u_z = 0$, $u_y = c$). (a) Write down the components of its velocity \mathbf{u}' relative to a frame S' traveling in the standard configuration with speed v along the x axis of frame S . (b) In what direction is the signal traveling relative to S' ? (c) Using your answer to part (a), calculate the magnitude of \mathbf{u}' .

SECTION 1.14 (The Doppler Effect*)

- 1.49** • It is found that the light from a nearby star is blueshifted by 1%; that is, $f_{\text{obs}} = 1.01 f_{\text{sec}}$. Is the star receding or approaching, and how fast is it traveling? (Assume that it is moving directly toward or away from us.)

- 1.50** • A star is receding from us at $0.5c$. What is the percent shift in the frequency of light received from the star?

- 1.51** • Consider the tale of the physicist who is ticketed for running a red light and argues that, because he was approaching the intersection, the red light was Doppler

shifted and appeared green. How fast would he have been going? ($\lambda_{\text{red}} \approx 650 \text{ nm}$, $\lambda_{\text{green}} \approx 530 \text{ nm}$.)

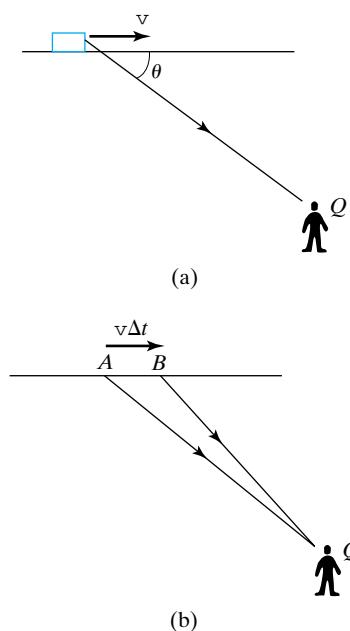
- 1.52** • In our discussion of the Doppler shift, we found three superficially different expressions for the received frequency f_{obs} , namely (1.51), (1.53), and (1.57). Show in detail that all three are equal.

- 1.53** •• Consider a source of light of frequency f_{sec} moving obliquely to an observer Q as in Fig. 1.24(a). (a) Prove that Q receives the light with frequency f_{obs} given by the general Doppler formula

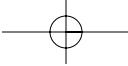
$$f_{\text{obs}} = \frac{f_{\text{sec}}}{(1 - \beta \cos \theta)\gamma}$$

- (b) Check that this formula reduces to our previous result (1.51) when the source is approaching Q head-on.

The analysis in part (a) is quite similar to that leading to (1.51) but the geometry is more complicated. Consider two successive wave crests emitted at points A and B as in Fig. 1.24(b). Since A and B are in practice very close together, the rays AQ and BQ are effectively parallel. Show that the difference between the lengths AQ and BQ is approximately $v \Delta t \cos \theta$ and hence that the distance between successive crests as they approach Q is $(c - v \cos \theta) \Delta t$. This is the appropriate generalization of (1.47), and from here the discussion is closely parallel.

**FIGURE 1.24**

(Problem 1.53) (a) Light from the moving source to the observer Q makes an angle θ with the velocity \mathbf{v} . (b) If two successive wave crests are emitted at A and B , a time Δt apart, then AB is $v \Delta t$.



Chapter 2

Relativistic Mechanics

- 2.1** Introduction
 - 2.2** Mass in Relativity
 - 2.3** Relativistic Momentum
 - 2.4** Relativistic Energy
 - 2.5** Two Useful Relations
 - 2.6** Conversion of Mass to Energy
 - 2.7** Force in Relativity
 - 2.8** Massless Particles
 - 2.9** When Is Nonrelativistic Mechanics Good Enough?
 - 2.10** General Relativity*
 - 2.11** The Global Positioning System: An Application of Relativity*
- Problems for Chapter 2

*These sections can be omitted without serious loss of continuity.

2.1 Introduction

We have seen that the laws of classical mechanics hold in a family of inertial frames that are related to one another by the classical, Galilean transformation; in other words, classical mechanics is invariant under the Galilean transformation. But we now know that the correct, relativistic transformation between inertial frames is the Lorentz, not the Galilean, transformation. It follows that the laws of classical mechanics cannot be correct and that we must find a new, relativistic mechanics that is invariant as we pass from one inertial frame to another, using the correct Lorentz transformation.

We will find that relativistic mechanics, like classical mechanics, is built around the concepts of mass, momentum, energy, and force. However, the relativistic definitions of these concepts are all a little different from their classical counterparts. In seeking these new definitions and the laws that connect them, we must be guided by three principles: First, a correct relativistic law must be valid in all inertial frames; it must be invariant under the Lorentz transformation. Second, we would expect the relativistic definitions and laws to reduce to their nonrelativistic counterparts when applied to systems moving much slower than the speed of light. Third, and most important, our relativistic laws must agree with experiment.

2.2 Mass in Relativity

We start by considering the mass m of an object — an electron, a space rocket, or a star. The most satisfactory definition of m turns out to be remarkably simple. We know that at slow speeds a suitable definition is the classical one (for

example, $m = F/a$, where a is the acceleration produced by a standard force F). In relativity we simply agree to use the same, classical definition of m with the proviso that *before measuring any mass we bring the object concerned to rest*. To emphasize this qualification, we will sometimes refer to m as the **rest mass**. It can also be called the **proper mass** since it is the mass measured in the frame where the object is at rest.

Observers in different inertial frames all agree on the rest mass of an object. Suppose that observers in a frame S take some object, bring it to rest (in S), and measure its mass m . If they then pass the object to observers in a different frame S' , who bring it to rest in S' and measure its mass m' , it will be found that $m' = m$. That is, rest mass is *invariant* as we pass from S to S' . In fact, this is required by the postulates of relativity: If m' were different from m , we could define a preferred frame (the frame where the rest mass of an object was minimum, for example).

As we will describe shortly, some physicists use a different definition of mass, called the *variable mass*. We will not use this concept, and whenever we use the word “mass” without qualification, we will mean the invariant rest mass defined here.

2.3 Relativistic Momentum

The classical definition of the momentum of a single body is

$$\mathbf{p} = m\mathbf{u} \quad (2.1)$$

where m and \mathbf{u} are the body’s mass and velocity. Since we know how both m and \mathbf{u} are measured in relativity, it is natural to start by asking the question: Is it perhaps the case that the classical definition (2.1) is the correct definition in relativity as well?

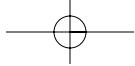
Strictly speaking, this question has no answer. There can be no such thing as a “correct” or “incorrect” definition of \mathbf{p} since one is at liberty to *define* things however one pleases. The proper question is rather: Is the definition $\mathbf{p} = m\mathbf{u}$ a *useful* definition in relativity?

In classical mechanics the concept of momentum has many uses, but its single most useful property is probably the law of conservation of momentum. If we consider n bodies with momenta $\mathbf{p}_1, \dots, \mathbf{p}_n$ then, in the absence of external forces, the total momentum

$$\sum \mathbf{p} = \mathbf{p}_1 + \cdots + \mathbf{p}_n$$

can never change. It would certainly be useful if we could find a definition of momentum such that this important law carried over into relativity. Accordingly, we will seek a definition of relativistic momentum \mathbf{p} with the following two properties:

1. $\mathbf{p} \approx m\mathbf{u}$ when $u \ll c$
2. The total momentum $\sum \mathbf{p}$ of an isolated system is conserved, as measured in every inertial frame.



48 Chapter 2 • Relativistic Mechanics

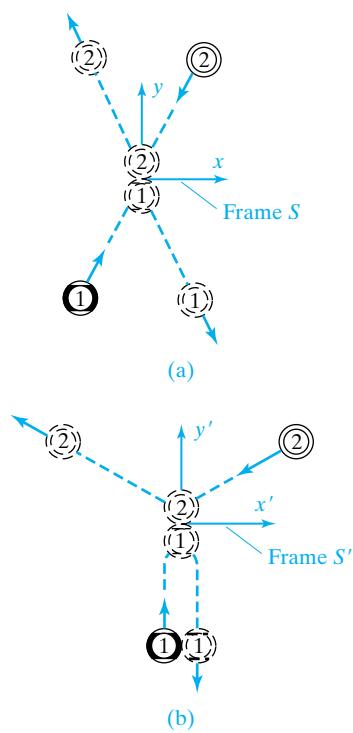


FIGURE 2.1

Two different views of a collision between two identical balls. (a) In frame S the velocities of the two balls are equal and opposite before and after the impact. (b) The same experiment as seen in frame S' , which travels along the x axis at the same rate as ball 1. In this frame ball 1 travels straight up the y' axis and back down again.

Requirement (1) is just that the relativistic definition should agree with the non-relativistic one in the nonrelativistic domain. Requirement (2) is that the law of conservation of momentum must hold in all inertial frames if it holds at all.

If we were to adopt the classical definition $\mathbf{p} = m\mathbf{u}$, requirement (1) would be satisfied automatically. However, it is fairly easy to construct a thought experiment which illustrates that the classical definition $\mathbf{p} = m\mathbf{u}$ does not meet requirement (2).

Consider two identical billiard balls that collide as shown in Fig. 2.1. Relative to the frame S of Fig. 2.1(a), the initial velocities of the two balls are equal and opposite. Further, the collision is arranged symmetrically, such that the x components of the velocities are unchanged by the impact, whereas the y components reverse. It is an experimental fact that such collisions between particles of equal mass do occur. It is easy to check that if we adopt the classical definition, $\mathbf{p} = m\mathbf{u}$, then as seen in S this collision conserves momentum,

$$\sum m\mathbf{u}(\text{before}) = \sum m\mathbf{u}(\text{after})$$

The proof that this is so is shown in Table 2.1, where we have denoted by a and b the x and y components of the initial velocity of ball 1. Notice that, as seen in S , the total classical momentum is actually zero before and after the collision; so it is certainly conserved.

TABLE 2.1

The experiment of Fig. 2.1 as seen in S . Each pair of numbers represents the x and y components of the vector indicated.

	First Ball (\mathbf{u}_1)	Second Ball (\mathbf{u}_2)	$m\mathbf{u}_1 + m\mathbf{u}_2$
Before:	(a, b)	$(-a, -b)$	$(0, 0)$
After:	$(a, -b)$	$(-a, b)$	$(0, 0)$

Let us now consider the same collision from a second frame S' traveling in the positive x direction of S , at the same rate as ball 1 (that is, the speed v of S' relative to S is equal to the x component of the velocity of ball 1). Figure 2.1(b) shows the collision as seen in S' , with ball 1 traveling straight up the y' axis and bouncing straight down again. If the frames S and S' were related by the classical, Galilean transformation (which we know is actually incorrect), the velocities measured in S' could be found from those in Table 2.1 using the classical velocity-addition formula, with the results shown in Table 2.2. From the last column, we see that the total classical momentum has the same values before and after impact; that is, the total classical momentum is

TABLE 2.2

The experiment of Fig. 2.1 as seen in frame S' , assuming that S and S' are related by the Galilean transformation. Each velocity here is obtained from that of Table 2.1 by subtracting a from its x component.

	First Ball (\mathbf{u}'_1)	Second Ball (\mathbf{u}'_2)	$m\mathbf{u}'_1 + m\mathbf{u}'_2$
Before:	$(0, b)$	$(-2a, -b)$	$(-2ma, 0)$
After:	$(0, -b)$	$(-2a, b)$	$(-2ma, 0)$

conserved in the frame S' also. This is just what we should have expected since we know that the laws of classical mechanics are invariant under the Galilean transformation.

In fact, however, the Galilean transformation is *not* the correct transformation between frames S and S' , and we must compute the velocities in S' using the relativistic transformations (1.43) and (1.44). The results of this rather tedious calculation (Problem 2.5) are shown in Table 2.3. The details of this table are not especially interesting, but there are two important points: First, because the transformation of u_y depends on u_x , the y components of the two balls' velocities transform differently, and, as seen in S' , they are no longer equal in magnitude. (Compare the y components in the first and second columns.) Consequently, the y component of the total classical momentum (final column of Table 2.3) is positive before the collision and negative after. Thus, even though the total classical momentum is conserved in frame S , it is *not* conserved in frame S' . Therefore, the law of conservation of classical momentum (defined as $\mathbf{p} = m\mathbf{u}$ for each body) is incompatible with the postulates of relativity, and the classical definition of momentum does not satisfy our requirement (2) listed above.

TABLE 2.3

The experiment of Fig. 2.1 as seen in S' , based on the correct relativistic velocity transformation from S to S' . The relative velocity of the two frames is $v = a$; accordingly, β denotes a/c and $\gamma = (1 - a^2/c^2)^{-1/2}$.

	First Ball (\mathbf{u}'_1)	Second Ball (\mathbf{u}'_2)	$m\mathbf{u}'_1 + m\mathbf{u}'_2$
Before:	$\left(0, \frac{b}{\gamma(1 - \beta^2)}\right)$	$\left(\frac{-2a}{1 + \beta^2}, \frac{-b}{\gamma(1 + \beta^2)}\right)$	$\left(\frac{-2ma}{1 + \beta^2}, \frac{2mb\beta^2}{\gamma(1 - \beta^4)}\right)$
After:	$\left(0, \frac{-b}{\gamma(1 - \beta^2)}\right)$	$\left(\frac{-2a}{1 + \beta^2}, \frac{b}{\gamma(1 + \beta^2)}\right)$	$\left(\frac{-2ma}{1 + \beta^2}, \frac{-2mb\beta^2}{\gamma(1 - \beta^4)}\right)$

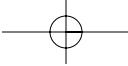
If there is a law of momentum conservation in relativity, the relativistic definition of momentum must be different from the classical one, $\mathbf{p} = m\mathbf{u}$. If we rewrite this classical definition as

$$\mathbf{p} = m \frac{d\mathbf{r}}{dt} \quad [\text{classical}] \quad (2.2)$$

we get a useful clue for a better definition. The difficulty in our thought experiment originated in the complicated transformation of the velocity $d\mathbf{r}/dt$ (particularly the y component). These complications arose because both $d\mathbf{r}$ and dt change as we transform from S to S' . We can avoid some of this problem if we replace the derivative $d\mathbf{r}/dt$ in (2.2) by the derivative $d\mathbf{r}/dt_0$ with respect to the *proper time*, t_0 , of the moving body:

$$\mathbf{p} = m \frac{d\mathbf{r}}{dt_0} \quad [\text{relativistic}] \quad (2.3)$$

In (2.2), $d\mathbf{r}$ is the vector joining two neighboring points on the body's path, and dt is the time for the body to move between them — both as measured in any one inertial frame S . In (2.3) $d\mathbf{r}$ is the same as before, but dt_0 is the *proper time*



50 Chapter 2 • Relativistic Mechanics

between the two points; that is, the time as measured in the body's rest frame. From its definition, the proper time dt_0 (just like the proper mass m) has the same value for all observers in all frames. Thus the vector defined in (2.3) transforms more simply than the classical momentum (2.2) since only the numerator $d\mathbf{r}$ changes as we move from one frame to another. In particular, the y component, $p_y = m dy/dt_0$, does not change at all as we pass from S to S' , and the difficulty encountered in our thought experiment would not occur if we were to adopt the definition (2.3). (For details, see Problem 2.6.)

At slow speeds, dt and dt_0 are indistinguishable and the new definition (2.3) agrees with the classical one; that is, the definition (2.3) meets requirement (1). Further, one can show that it always meets requirement (2); specifically, if the total momentum $\sum \mathbf{p}$, as defined by (2.3), is constant in one inertial frame S , the same is true in *all* inertial frames.* Since the proof is rather long, although reasonably straightforward, we leave it as a problem (Problem 2.14) at the end of this chapter.

With the definition (2.3) of momentum, the law of conservation of momentum would be logically consistent with the postulates of relativity. Whether momentum defined in this way is conserved must be decided by experiment. The unanimous verdict of innumerable experiments involving collisions of atomic and subatomic particles is that it is: If we adopt the definition (2.3) for the momentum of a body, the total momentum $\sum \mathbf{p}$ of an isolated system *is* conserved. Under the circumstances, we naturally adopt (2.3) as our definition of momentum.

It is convenient to express the definition (2.3) a little differently: The time-dilation formula implies that $dt = \gamma dt_0$, where $\gamma = (1 - u^2/c^2)^{-1/2}$. Therefore, (2.3) is the same as

$$\mathbf{p} = m \frac{d\mathbf{r}}{dt_0} = m\gamma \frac{d\mathbf{r}}{dt} = \gamma m\mathbf{u}$$

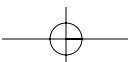
Thus we adopt as the final form of our definition:

The momentum of a single body of mass m and velocity \mathbf{u} is

$$\mathbf{p} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} = \gamma m\mathbf{u} \quad (2.4)$$

An important consequence of the factor γ in the momentum (2.4) is that no object can be accelerated past the speed of light: We will find that in relativity, just as in classical mechanics, a constant force on a body increases its momentum p at a constant rate; if the force acts for long enough, we can make p as large as we please. In classical mechanics, where $p = mu$, this means that a constant force steadily increases u and can eventually make u as large as we please. In relativity, an increase in $p = \gamma mu$ is reflected by increases in u and γ . Now, as $u \rightarrow c$ we know that γ increases without limit. Thus, as $u \rightarrow c$ the

*The proposition as stated is a little oversimplified. What one can actually prove is this: If both momentum, as defined by (2.3), *and* energy (whose definition we give in Section 2.4) are conserved in one inertial frame, they are both automatically conserved in all inertial frames.



constant force keeps increasing γ without u ever reaching c . This difference between classical and relativistic mechanics is illustrated in Fig. 2.2, which shows a plot of u against p for both cases.

Example 2.1

A 1-kg lump of metal is observed traveling with speed $0.4c$. What is its momentum? What would its momentum be if we doubled its speed? Compare with the corresponding classical values.

When $\beta = 0.4$, the factor $\gamma = (1 - \beta^2)^{-1/2}$ is easily calculated to be $\gamma = 1.09$, and

$$\begin{aligned} p &= \gamma mu = 1.09 \times (1 \text{ kg}) \times (0.4 \times 3 \times 10^8 \text{ m/s}) \\ &= 1.31 \times 10^8 \text{ kg} \cdot \text{m/s} \end{aligned} \quad (2.5)$$

If we double the speed, $\beta = 0.8$ and $\gamma = 1.67$. Thus the momentum becomes

$$\begin{aligned} p &= \gamma mu = 1.67 \times (1 \text{ kg}) \times (0.8 \times 3 \times 10^8 \text{ m/s}) \\ &= 4.01 \times 10^8 \text{ kg} \cdot \text{m/s} \end{aligned} \quad (2.6)$$

which is more than three times the previous answer. The classical answers are found by omitting the factors of γ : If $\beta = 0.4$, then $p = 1.20 \times 10^8 \text{ kg} \cdot \text{m/s}$, just a little less than the correct answer (2.5); if $\beta = 0.8$, then $p = 2.40 \times 10^8 \text{ kg} \cdot \text{m/s}$, significantly less than the correct answer (2.6).

Some physicists like to think of the relativistic momentum as the product of γm and \mathbf{u} , which they write as

$$\mathbf{p} = m_{\text{var}} \mathbf{u} \quad (2.7)$$

where

$$m_{\text{var}} = \gamma m = \frac{m}{\sqrt{1 - u^2/c^2}}. \quad (2.8)$$

The quantity m_{var} is called the *variable mass* since, unlike the rest mass m , it varies with the body's speed u . The form (2.7) has the advantage of making the relativistic momentum look more like its nonrelativistic counterpart $\mathbf{p} = m\mathbf{u}$. On the other hand, it is not always a good idea to give two ideas the appearance of similarity when they are in truth different. Further, the introduction of the variable mass does not achieve a complete parallel with classical mechanics. For example, we will see that the quantity $\frac{1}{2}m_{\text{var}}u^2$ is *not* the correct expression for the relativistic kinetic energy and the equation $\mathbf{F} = m_{\text{var}}\mathbf{a}$ is *not* the correct relativistic form of Newton's second law (Problems 2.13 and 2.35). For these reasons, we will not use the notion of variable mass in this book.

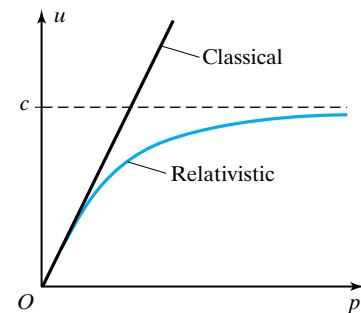
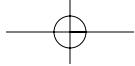


FIGURE 2.2

The speed of a body as a function of its momentum in classical and relativistic mechanics. At low speeds the two curves merge. In classical mechanics u grows indefinitely as p increases. In relativity u never exceeds c , however large p becomes; instead, u is asymptotic to c as $p \rightarrow \infty$.

2.4 Relativistic Energy

Having found a suitable relativistic definition for momentum \mathbf{p} , our next task is to do the same for the energy E of a body. Just as with momentum, one is in principle free to define E however one pleases. But the hope of finding a *useful* definition suggests two requirements analogous to those used for momentum:



52 Chapter 2 • Relativistic Mechanics

1. When applied to slowly moving bodies, the new definition of E should reproduce as closely as possible the classical definition.
2. The total energy $\sum E$ of an isolated system of bodies should be conserved in all inertial frames.

The definition that fits these requirements turns out to be this:

The energy of a single body of mass m , moving with speed u , is

$$E = \frac{mc^2}{\sqrt{1 - u^2/c^2}} = \gamma mc^2 \quad (2.9)$$

It is important to note that this applies to any single body — an elementary particle, like an electron; an assembly of particles, like an atom; or an assembly of atoms, like a baseball, a space rocket, or a star.

Although we will not do so here (but see Problem 2.14), one can prove that with the definition (2.9), a law of conservation of energy would be logically consistent with the postulates of relativity: If $\sum E$ were constant as measured in one inertial frame, the same would be true in all inertial frames. Furthermore, experiment shows that the quantity $\sum E$ is conserved for any isolated system. Thus the definition (2.9) meets requirement (2).

Just because the quantity $\sum E$ is conserved, we are not yet justified in giving E , as defined by (2.9), the name *energy*. The main reason for doing so will emerge when we establish the connection of (2.9) with the classical definition of energy, that is, when we check requirement (1). Before we do so, we mention two other important points. First, since γ is dimensionless and mc^2 has the dimensions of energy, our definition $E = \gamma mc^2$ at least has the correct dimensions for an energy. Second, although we have not yet defined the concept of force in relativity, when we do so in Section 2.7, we will prove the following important theorem: If a total force \mathbf{F} acts on a body as it moves through a small displacement $d\mathbf{r}$, the resulting change in the energy E , as defined by (2.9), is

$$dE = \mathbf{F} \cdot d\mathbf{r} \quad (2.10)$$

You should recognize the product on the right as the work done by the force \mathbf{F} , and the equation (2.10) as the work-energy theorem: The change in a body's energy is the work done on it. The fact that this theorem applies to our new definition of E is strong reason for regarding E as the relativistic generalization of the classical notion of energy.

Let us now evaluate the relativistic energy (2.9) for a slowly moving body. With $u \ll c$, we can use the binomial approximation to write the factor γ as

$$\gamma = \left(1 - \frac{u^2}{c^2}\right)^{-1/2} \approx 1 + \frac{1}{2} \frac{u^2}{c^2}$$

Therefore,

$$E = \gamma mc^2 \approx \left(1 + \frac{1}{2} \frac{u^2}{c^2}\right) mc^2 = mc^2 + \frac{1}{2} mu^2 \quad [\text{when } u \ll c] \quad (2.11)$$

We see that for a slowly moving body, the relativistic energy is the sum of two terms: a constant term mc^2 that is independent of u and a second term $\frac{1}{2}mu^2$ that is precisely the classical kinetic energy of a body of mass m and speed u .

In classical physics it was believed that mass was always conserved, and the term mc^2 in (2.11) would therefore have been an immutable constant. Further, you will recall that one was always at liberty to add or subtract an overall constant from the energy since the zero of energy was arbitrary. Thus, in the classical context, (2.11) implies that when $u \ll c$ the relativistic energy E of a body is just the classical kinetic energy plus an irrelevant constant mc^2 . Therefore, the relativistic definition (2.9) meets both our requirements (1) and (2), and our identification of (2.9) as the appropriate generalization of the classical notion of energy is complete.

We will find that in relativity the “irrelevant constant” mc^2 in (2.11) is actually extremely important. The reason is that the classical law of conservation of mass turns out to be wrong. This law was based mainly on nineteenth-century measurements of masses in chemical reactions, where no change of mass was ever detected. In this century, however, nuclear processes have been discovered in which large changes of mass occur, and even where the rest mass of certain particles disappears entirely. Further, we now know that even in chemical reactions the total rest mass of the participating atoms and molecules *does* change, although the changes are much too small (1 part in 10^9 or so) to be detected directly.

Given that rest masses *can* change, it should be clear that the term mc^2 in (2.11) is important. In Section 2.6 we describe some processes in which the rest mass of a system does change, and will see just how important the term mc^2 is. In the remainder of this section we give two more definitions connected with the relativistic energy (2.9) and describe an application of the laws of energy and momentum conservation to processes in which rest masses do not change — the so-called *elastic* processes.

It is clear from either the exact equation $E = \gamma mc^2$ or the nonrelativistic approximation (2.11) that even when a body is at rest, its energy is not zero but is given, instead, by the famous equation

$$E = mc^2 \quad [\text{when } u = 0] \quad (2.12)$$

This energy is called the **rest energy** of the mass m , and we will see in Section 2.6 how it can be converted into other forms, such as the kinetic energy of other bodies.

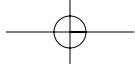
Example 2.2

What is the rest energy of a 1-kg lump of metal?

Substituting into (2.12), we find that

$$E = mc^2 = (1 \text{ kg}) \times (3 \times 10^8 \text{ m/s})^2 = 9 \times 10^{16} \text{ joules}$$

about the energy generated by a large power plant in one year. This incredible amount of energy would be of no interest if it could not be converted into other forms of energy. In fact, however, such conversion is possible. For example, if the metal is uranium 235, about 1 part in 1000 of the rest energy can be converted into heat by the process of nuclear fission. Thus 1 kg of ^{235}U can yield a fantastic 9×10^{13} joules of heat.



54 Chapter 2 • Relativistic Mechanics

When a body is not at rest, we can think of its total energy $E = \gamma mc^2$ as the sum of its rest energy mc^2 plus the additional energy $(E - mc^2)$ that it has by virtue of its motion. This second term we naturally call the **kinetic energy** K , and we write

$$E = mc^2 + K$$

where

$$K = E - mc^2 = (\gamma - 1)mc^2 \quad (2.13)$$

At slow speeds we have seen that $K \approx \frac{1}{2}mu^2$, but in general the relativistic kinetic energy is different from $\frac{1}{2}mu^2$. In particular, as $u \rightarrow c$, the kinetic energy approaches infinity. However much energy we give a body, its speed can therefore never reach the speed of light — a conclusion we reached before by considering the relativistic momentum. Notice that, since $\gamma \geq 1$, the relativistic kinetic energy is always positive (like its nonrelativistic counterpart $\frac{1}{2}mu^2$).

In classical mechanics a surprising number of interesting problems can be solved using just the laws of energy and momentum conservation. In relativity also, there are many such problems, and we conclude this section with an example of one.

Example 2.3

Two particles with rest masses m_1 and m_2 collide head-on, as shown in Fig. 2.3. Particle 1 has initial velocity u_1 , while particle 2 is at rest ($u_2 = 0$). Assuming that the collision is elastic (that is, the rest masses are unchanged), use conservation of energy and momentum to find the velocity u_3 of particle 1 after the collision. Apply the result to the case that a pion (a subatomic particle with mass $m_1 = 2.49 \times 10^{-28}$ kg) traveling at $0.9c$ makes an elastic head-on collision with a stationary proton ($m_2 = 1.67 \times 10^{-27}$ kg).

Since the solution of this problem is very similar to that of the corresponding nonrelativistic problem, let us first review the latter. Conservation of energy implies that

$$E_1 + E_2 = E_3 + E_4$$

If the rest masses of the two particles are unchanged, this can be rewritten as

$$(m_1c^2 + K_1) + (m_2c^2 + K_2) = (m_1c^2 + K_3) + (m_2c^2 + K_4)$$

and, canceling the mass terms, we see that kinetic energy is conserved,

$$K_1 + K_2 = K_3 + K_4$$

which is the usual definition of an elastic collision in classical mechanics. The conservation of momentum and kinetic energy imply (in nonrelativistic mechanics)

$$m_1u_1 = m_1u_3 + m_2u_4 \quad (2.14)$$



FIGURE 2.3

An elastic, head-on collision.

and

$$\frac{1}{2}m_1u_1^2 = \frac{1}{2}m_1u_3^2 + \frac{1}{2}m_2u_4^2. \quad (2.15)$$

These two equations can be solved for the two unknowns u_3 and u_4 . Since one of the equations is quadratic, we get two solutions. The first solution, $u_3 = u_1$ and $u_4 = 0$, gives the initial velocities before the collision occurred. The second solution is the interesting one and gives (Problem 2.11)

$$u_3 = \frac{m_1 - m_2}{m_1 + m_2}u_1 \quad (2.16)$$

and

$$u_4 = \frac{2m_1}{m_1 + m_2}u_1 \quad (2.17)$$

Several features of these answers deserve comment. If $m_1 > m_2$, u_3 is positive and particle 1 continues in its original direction; if $m_1 < m_2$, u_3 is negative and particle 1 bounces back in the opposite direction. If $m_1 = m_2$, $u_3 = 0$ and $u_4 = u_1$; that is, particle 1 comes to a dead stop, giving all of its momentum and kinetic energy to particle 2. If $m_1 \ll m_2$ (1 is much lighter than 2), $u_3 \approx -u_1$ and particle 1 bounces back off the much heavier target with its speed barely changed.

The solution of the corresponding relativistic problem is very similar, although considerably messier because of the square roots involved in the factors γ . Because there are several different velocities in the problem, we must be careful to distinguish the corresponding factors of γ . We therefore write γ_1 for $(1 - u_1^2/c^2)^{-1/2}$ and γ_2 for $(1 - u_2^2/c^2)^{-1/2}$ (which is equal to 1 in this problem), and so on. The conservation of relativistic momentum implies that

$$\gamma_1 m_1 u_1 = \gamma_3 m_1 u_3 + \gamma_4 m_2 u_4$$

while conservation of relativistic energy implies that

$$\gamma_1 m_1 c^2 + m_2 c^2 = \gamma_3 m_1 c^2 + \gamma_4 m_2 c^2 \quad (2.18)$$

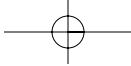
These are two equations for the two unknowns, u_3 and u_4 . If, for example, we eliminate u_4 , some fairly messy algebra leads us to a quadratic equation for u_3 . One solution of this equation is $u_3 = u_1$ (the original velocity before the collision) and the other is

$$u_3 = \frac{m_1^2 - m_2^2}{m_1^2 + m_2^2 + 2m_1 m_2 \sqrt{1 - u_1^2/c^2}} u_1 \quad (2.19)$$

The answer (2.19) has much in common with the nonrelativistic (2.16). When $m_1 > m_2$, particle 1 continues in its original direction (u_3 positive); if $m_1 < m_2$, particle 1 bounces back (u_3 negative). If $m_1 = m_2$, particle 1 comes to a dead stop. If particle 1 is moving nonrelativistically, the square root in the denominator can be replaced by 1 to give

$$u_3 \approx \frac{m_1^2 - m_2^2}{(m_1 + m_2)^2} u_1 = \frac{m_1 - m_2}{m_1 + m_2} u_1$$

which is precisely the nonrelativistic answer (2.16).



56 Chapter 2 • Relativistic Mechanics

For the case of a pion, with $u_1 = 0.9c$, colliding with a stationary proton we can substitute the given numbers into (2.19) to give

$$u_3 = -0.76c \quad [\text{relativistic}]$$

(Notice that this is negative, indicating that the light pion bounces backwards.) If we were to put the same numbers into the nonrelativistic result (2.16), we would find that

$$u_3 = -0.62c \quad [\text{nonrelativistic}]$$

The difference between these two answers is large enough to be easily detected, and it is the relativistic answer that proves to be correct. More generally, in all collisions involving atomic and subatomic particles, one finds perfect agreement between the experimental observations and the predictions based on conservation of relativistic energy and momentum. This is, in fact, the principal evidence that these quantities are conserved.

2.5 Two Useful Relations

We have introduced four parameters, m , \mathbf{u} , \mathbf{p} , and E , that characterize the motion of a body. Only two of these are independent since \mathbf{p} and E were defined in terms of m and \mathbf{u} as

$$\mathbf{p} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} \quad (2.20)$$

and

$$E = \frac{mc^2}{\sqrt{1 - u^2/c^2}} \quad (2.21)$$

We can, of course, rearrange these definitions to give an expression for any one of our parameters in terms of any two others. There are two such expressions that are especially useful, as we discuss now.

First, dividing (2.20) by (2.21), we find that

$$\frac{\mathbf{p}}{E} = \frac{\mathbf{u}}{c^2}$$

or

$$\beta \equiv \frac{\mathbf{u}}{c} = \frac{\mathbf{p}c}{E} \quad (2.22)$$

which gives the dimensionless “velocity” $\beta = \mathbf{u}/c$ in terms of \mathbf{p} and E . (Since $\mathbf{p}c$ has the dimensions of energy, the right side is dimensionless, as it must be.)

Second, by squaring both (2.20) and (2.21) it is easy to verify that (Problem 2.16)

$$E^2 = (pc)^2 + (mc^2)^2 \quad (2.23)$$

This useful expression for E in terms of p and m shows that the three quantities E , pc , and mc^2 are related like the sides of a right-angled triangle, with E as the hypotenuse, as illustrated in Fig. 2.4. [At this stage there is no deep geometrical significance to this statement; we mention it only as a convenient way to remember the relation (2.23).] In our applications of relativistic mechanics, we will frequently use both the result (2.22) and the “**Pythagorean relation**” (2.23).

Now is a convenient point to mention some of the units used to measure the parameters u , E , m , and p . First, as we have seen repeatedly, relativistic velocities are best expressed as fractions of c or by using the dimensionless $\beta = u/c$. For energy, the SI unit is the joule (J). However, most of our applications of relativity will be in atomic and subatomic physics, where the joule is an inconveniently large unit of energy, and a much more popular unit is the **electron volt** or eV. This is defined as the work needed to move an electron (of charge $q = -e = -1.60 \times 10^{-19}$ coulomb) through a voltage drop of 1 volt ($\Delta V = -1$ volt); thus

$$\begin{aligned} 1 \text{ eV} &= q \Delta V = (-1.60 \times 10^{-19} \text{ coulomb}) \times (-1 \text{ volt}) \\ &= 1.60 \times 10^{-19} \text{ J}. \end{aligned} \quad (2.24)$$

We shall find that typical energies in atomic physics are of the order of 1 eV or so; those in nuclear physics are of order 10^6 eV, or 1 MeV.

Like the joule, the kilogram is inconveniently large as a unit for atomic and subatomic physics. For example, the mass of the electron is 9.11×10^{-31} kg. In fact, in most applications of relativity one is not so much concerned with the mass m as with the corresponding rest energy mc^2 . [For instance, in (2.23) it is mc^2 , rather than m itself, that appears.] Thus in many relativistic problems, masses are given implicitly by stating the rest energy mc^2 , usually in eV. Another way to say this is that mass is measured in units of eV/c^2 , as we discuss in the following example.

Example 2.4

Given that the electron has mass $m = 9.109 \times 10^{-31}$ kg, what is its rest energy in eV, and what is its mass in eV/c^2 ?

The rest energy is

$$\begin{aligned} mc^2 &= (9.109 \times 10^{-31} \text{ kg}) \times (2.998 \times 10^8 \text{ m/s})^2 \\ &= (81.87 \times 10^{-15} \text{ J}) \times \frac{1 \text{ eV}}{1.602 \times 10^{-19} \text{ J}} \\ &= 5.11 \times 10^5 \text{ eV} \end{aligned}$$

or*

$$mc^2 = 0.511 \text{ MeV} \quad (2.25)$$

*Note that to get this answer, which is correct to three significant figures, we used four significant figures in all input numbers. These were taken from Appendix A, which lists the best known values of the fundamental constants.

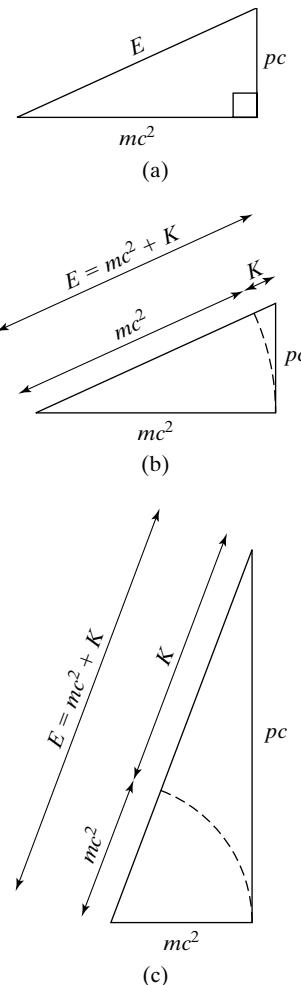
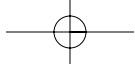


FIGURE 2.4

- (a) The “Pythagorean relation” (2.23) means that the three variables E , pc , and mc^2 form the sides of a right triangle, with E as the hypotenuse. (b) Since $E = mc^2 + K$, the hypotenuse can be divided as shown. When pc is much smaller than mc^2 , the energy is mostly rest energy. (c) When pc is larger than mc^2 , the energy is mostly kinetic.



58 Chapter 2 • Relativistic Mechanics

Thus a convenient way to specify (and remember) the electron's mass is to say that its rest energy mc^2 is roughly half an MeV.

An alternative way to put this is to divide both sides of (2.25) by c^2 and say that

$$m = 0.511 \text{ MeV}/c^2$$

If you have not met eV/c^2 before as a unit of mass, it will probably seem a bit odd at first. The important thing to remember is that the statement " $m = 0.5 \text{ MeV}/c^2$ " is precisely equivalent to the more transparent " $mc^2 = 0.5 \text{ MeV}$ ".

In most applications of relativity we will be less interested in the momentum p than in the product pc . [For example, both of the relations (2.22) and (2.23) involve pc rather than p .] The quantity pc has the dimensions of energy and so is often measured in eV or MeV. This is equivalent to measuring p itself in eV/c or MeV/c .

Example 2.5

An electron (rest mass about $0.5 \text{ MeV}/c^2$) is moving with total energy $E = 1.3 \text{ MeV}$. Find its momentum (in MeV/c , and in SI units) and its speed.

Given the energy and mass, we can immediately find the momentum from the "Pythagorean relation" (2.23)

$$\begin{aligned} pc &= \sqrt{E^2 - (mc^2)^2} = \sqrt{(1.3 \text{ MeV})^2 - (0.5 \text{ MeV})^2} \\ &= 1.2 \text{ MeV} \end{aligned}$$

or

$$p = 1.2 \text{ MeV}/c$$

(Notice how easily the units work out when we measure E in MeV, p in MeV/c , and m in MeV/c^2 .) This is easily converted into SI units if necessary: The required conversion is

$$\begin{aligned} 1 \frac{\text{MeV}}{c} &= \frac{1.602 \times 10^{-13} \text{ J}}{2.998 \times 10^8 \text{ m/s}} \\ &= 5.34 \times 10^{-22} \text{ kg} \cdot \text{m/s} \end{aligned} \tag{2.26}$$

(This and several other conversion factors are listed inside the front cover; more exact values are given in Appendix A.) Therefore,

$$\begin{aligned} p &= (1.2 \text{ MeV}/c) \times \frac{5.3 \times 10^{-22} \text{ kg} \cdot \text{m/s}}{1 \text{ MeV}/c} \\ &\approx 6 \times 10^{-22} \text{ kg} \cdot \text{m/s}. \end{aligned}$$

Once we know the energy and momentum, the speed follows immediately from (2.22).

$$\beta = \frac{pc}{E} = \frac{1.2 \text{ MeV}}{1.3 \text{ MeV}} = 0.92 \quad \text{or} \quad u \approx 0.9c$$

2.6 Conversion of Mass to Energy

According to the relativistic definition of energy, even a mass at rest has an energy, equal to mc^2 . As we have emphasized, this statement is meaningless unless there is some way in which the supposed rest energy mc^2 can be converted, or at least partly converted, into other more familiar forms of energy, such as kinetic energy — a process that is often loosely called conversion of mass to energy. Such conversion would require the classical law of conservation of mass to be violated. In this section we argue that if the relativistic mechanics we have developed is correct, the nonconservation of mass is logically necessary; and we describe some processes in which mass is not conserved.

Let us consider two bodies that can come together to form a single composite body. The two bodies could be two atoms that can bind together to form a molecule, or two atomic nuclei that can fuse to form a larger nucleus. Because it is probably easier to think about everyday objects, we will imagine two macroscopic (that is, nonmicroscopic) blocks. When the bodies are far apart, we take for granted that the forces between them are negligible. But when they are close together, there will be forces, and we distinguish two cases: If the forces are predominantly repulsive, we will have to supply energy to push the two bodies together; if the forces are predominantly attractive, energy will be released as they move together.

As a model of the repulsive case, we imagine two blocks with a compressible spring attached to one of them, as shown in Fig. 2.5(a). To hold the blocks together, we attach a pivoted catch to the second block, as shown. Suppose now that we push the two blocks together until the catch closes, as in Fig. 2.5(b). This will require us to do work, which is stored by the system as potential energy of the spring. Because of this potential energy, the bound state of the system is unstable: If we release the catch, the blocks will fly apart and the stored energy will reappear as kinetic energy, as shown in Fig. 2.5(c). What we will now argue is that the potential energy stored in the bound system manifests itself as an increase in the mass of the system; that is, the rest mass M of the unstable bound state is greater than the sum of the separate rest masses m_1 and m_2 of the blocks when far apart. (In classical mechanics, $M = m_1 + m_2$ of course.)

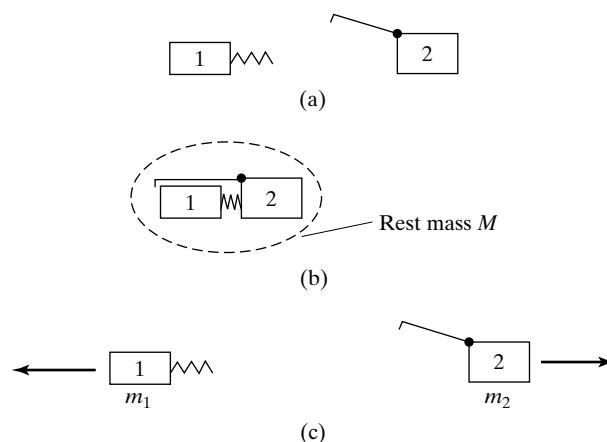
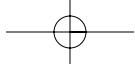


FIGURE 2.5

(a) Two blocks repel one another at close range because of the spring attached to block 1. (b) If the blocks are pushed together, the catch on block 2 can hold them in a bound state of rest mass M . (c) The bound state is unstable in the sense that the blocks fly apart when the catch is released.



60 Chapter 2 • Relativistic Mechanics

As long as the two blocks remain together, we can treat them as a single body, whose rest mass we denote by M . If the body is at rest, then according to the definition (2.9) of relativistic energy its total energy is

$$\text{total energy} = Mc^2 \quad (2.27)$$

[Notice that we are assuming that the definition (2.9) can be applied to a composite system like our two blocks locked together; in the final analysis this assumption must be — and is — justified by experiment.] If we now gently release the catch, the two blocks will fly apart. Once they are well separated, we can treat them as two separate bodies with rest masses m_1 and m_2 and total energy.

$$\begin{aligned} \text{total energy} &= E_1 + E_2 \\ &= K_1 + m_1 c^2 + K_2 + m_2 c^2 \end{aligned} \quad (2.28)$$

By conservation of energy, (2.27) and (2.28) must be equal. Therefore,

$$Mc^2 = (K_1 + K_2) + (m_1 + m_2)c^2 \quad (2.29)$$

Since both K_1 and K_2 are greater than zero, we are forced to the conclusion that M is greater than the sum of the separate masses, $m_1 + m_2$. This conclusion is independent of the mechanical details of our example and applies to any unstable system of rest mass M that can fly apart into two (or more) pieces.

We can solve (2.29) to give the amount by which the rest mass decreases as the blocks fly apart. If we denote this by

$$\Delta M = M - (m_1 + m_2)$$

then (2.29) implies that

$$\begin{aligned} \Delta M c^2 &= K_1 + K_2 \\ &= \text{energy released as bodies fly apart} \end{aligned} \quad (2.30)$$

We can say that a mass ΔM has been converted into kinetic energy, the rate of exchange between mass and energy being given by the familiar “energy = mass $\times c^2$.”

The kinetic energy released as the bodies move apart is the same as the work done to bring them together in the first place. Thus we can rephrase our result to say that when we push our two blocks together, the work done results in an increase in rest mass, ΔM , given by

$$\Delta M c^2 = \text{work done to push bodies together} \quad (2.31)$$

Example 2.6

The nuclei of certain atoms are naturally unstable, or radioactive, and spontaneously fly apart, tearing the whole atom into two pieces. For example, the atom called thorium 232 splits spontaneously into two “offspring” atoms, radium 228 and helium 4,



Section 2.6 • Conversion of Mass to Energy 61

The combined kinetic energy of the two offspring is 4 MeV. By how much should the rest mass of the “parent” ^{232}Th differ from the combined rest mass of its offspring? Compare this with the difference in the measured masses listed in Appendix D.

The reaction (2.32) is analogous to the example of the two blocks just discussed. (In fact, the analogy is remarkably good. The repulsive force is the electrostatic repulsion of the offspring nuclei, both of which are positively charged; the “catch” that holds the offspring together is their nuclear attraction, which is not quite strong enough and eventually “releases,” letting the offspring fly apart.) The required mass difference ΔM is given by (2.30) as

$$\Delta M = \frac{K_1 + K_2}{c^2} = 4 \text{ MeV}/c^2$$

Therefore, ^{232}Th should be heavier than the combined mass of ^{228}Ra and ^4He by $4 \text{ MeV}/c^2$. We can convert this to kilograms if we wish.

$$\begin{aligned}\Delta M &= 4 \frac{\text{MeV}}{c^2} = 4 \times \frac{10^6 \times 1.6 \times 10^{-19} \text{ J}}{(3 \times 10^8 \text{ m/s})^2} \\ &= 7 \times 10^{-30} \text{ kg}\end{aligned}\quad (2.33)$$

This mass difference is very small, even compared to the masses of the atoms involved. (^{232}Th has a mass of about 4×10^{-25} kg; so the mass difference is of order 1 part in 10^5 of the total mass.) Nevertheless, it is large enough to be measured directly.

To check our predicted mass difference against the measured masses, we refer to Appendix D, which lists the masses concerned as follows:

Atom	Mass (in u)		
^{232}Th	232.038	Initial	
^{228}Ra	228.031		
^4He	+ 4.003 = 232.034 0.004	Final Total Difference	

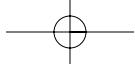
These masses are given in atomic mass units (denoted u). We will discuss this unit in Chapter 3, but for now we need to know only that

$$1 \text{ u} = 1.66 \times 10^{-27} \text{ kg}$$

Thus the measured mass difference is

$$\begin{aligned}\Delta M &= 0.004 \text{ u} \times \frac{1.7 \times 10^{-27} \text{ kg}}{1 \text{ u}} \\ &= 7 \times 10^{-30} \text{ kg}\end{aligned}$$

in agreement with our prediction (2.33).



62 Chapter 2 • Relativistic Mechanics

We next consider briefly the case that the force between the bodies is predominantly attractive, so that work is required to pull them apart. For example, if our two bodies are an electron and a proton, they attract one another because of their opposite electric charges; if they come close together, they can form the stable bound state that we call the hydrogen atom and an external agent must do work (13.6 eV, in fact) to pull them apart again. The work needed to pull a bound state apart (leaving the pieces well separated and at rest) is called the **binding energy** and is denoted by B . As before, we denote by M the rest mass of the bound state and by m_1 and m_2 the separate masses of the two bodies. By conservation of energy

$$Mc^2 + B = m_1c^2 + m_2c^2 \quad (2.34)$$

In this case we see that M is less than $m_1 + m_2$, the difference $\Delta M = m_1 + m_2 - M$ being given by

$$\begin{aligned} \Delta Mc^2 &= \text{binding energy, } B \\ &= \text{work to pull bodies apart} \end{aligned} \quad (2.35)$$

If we now released our two bodies from rest, they would accelerate back together and could reenter their bound state, with the release of energy B . (In the example of the electron and proton, the energy is released as light when they form into a hydrogen atom.) Thus we can rephrase (2.35) to say that as the two bodies come together and form the stable bound state, there is a release of energy and corresponding loss* of mass, ΔM , given by

$$\Delta Mc^2 = \text{energy released as bodies come together to form bound state} \quad (2.36)$$

Example 2.7

It is known that two oxygen atoms attract one another and can unite to form an O_2 molecule, with the release of energy $E_{\text{out}} \approx 5 \text{ eV}$ (in the form of light if the reaction takes place in isolation). By how much is the O_2 molecule lighter than two O atoms? If one formed 1 gram of O_2 in this way, what would be the total loss of rest mass and what is the total energy released? (The O_2 molecule has a mass of about $5.3 \times 10^{-26} \text{ kg}$.)

By (2.36) the mass of one O_2 molecule is less than that of two O atoms by an amount

$$\begin{aligned} \Delta M &= \frac{E_{\text{out}}}{c^2} \approx 5 \text{ eV}/c^2 \\ &\approx 5 \times \frac{1.6 \times 10^{-19} \text{ J}}{(3 \times 10^8 \text{ m/s})^2} \approx 9 \times 10^{-36} \text{ kg} \end{aligned}$$

*It is quite easy to confuse the direction of the mass change (gain or loss) in a given process. If this happens, just go back to the fact of *energy conservation*, which you can easily write in a form like (2.29) or (2.34). From this you can see immediately which mass is greater.

Section 2.6 • Conversion of Mass to Energy 63

Dividing this by the mass 5.3×10^{-26} kg of a single O₂ molecule, we see that the *fractional loss* of mass is about 2 parts in 10^{10} . Thus if we were to form 1 gram (g) of O₂ this way, the total loss of mass would be

$$\Delta M \approx 2 \times 10^{-10} \text{ g} \quad (2.37)$$

which is much too small to be measured directly. This is fairly typical of the mass changes in chemical reactions and explains why nonconservation of mass does not show up in chemistry.

While the mass change (2.37) is exceedingly small, the total energy released,

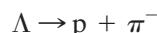
$$\begin{aligned} E_{\text{out}} &= \Delta M c^2 \approx (2 \times 10^{-13} \text{ kg}) \times (3 \times 10^8 \text{ m/s})^2 \\ &\approx 2 \times 10^4 \text{ J} \end{aligned}$$

is large — much more than would be needed to boil a gram of water, for example. This is because the conversion factor, c^2 , from mass to energy is so large.

So far in this section we have focused on the conservation of relativistic energy and the concomitant nonconservation of mass. For an isolated system, momentum is also conserved, and in many problems conservation of energy and momentum give enough information to determine all that one needs to know. We conclude this section with an example.

Example 2.8

The Λ particle is a subatomic particle that (as mentioned in Example 1.2) can decay spontaneously into a proton and a negatively charged pion.



(This immediately tells us that the rest mass of the Λ is greater than the total rest mass of the proton and pion.) In a certain experiment the outgoing proton and pion were observed, both traveling in the same direction along the positive x axis with momenta

$$p_p = 581 \text{ MeV}/c \quad \text{and} \quad p_\pi = 256 \text{ MeV}/c$$

Given that their rest masses are known to be

$$m_p = 938 \text{ MeV}/c^2 \quad \text{and} \quad m_\pi = 140 \text{ MeV}/c^2$$

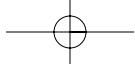
find the rest mass m_Λ of the Λ .

We can solve this problem in three steps: First, knowing p and m for the proton and pion, we can calculate their energies using the “Pythagorean relation”

$$E^2 = (pc)^2 + (mc^2)^2 \quad (2.38)$$

This gives

$$E_p = 1103 \text{ MeV} \quad \text{and} \quad E_\pi = 292 \text{ MeV}$$



64 Chapter 2 • Relativistic Mechanics

Second, using conservation of energy and momentum, we can reconstruct the energy and momentum of the original Λ :

$$E_\Lambda = E_p + E_\pi = 1395 \text{ MeV} \quad (2.39)$$

and

$$\mathbf{p}_\Lambda = \mathbf{p}_p + \mathbf{p}_\pi = 837 \text{ MeV}/c \quad (\text{along the positive } x \text{ axis}) \quad (2.40)$$

Finally, knowing E and p for the Λ , we can use the relation (2.38) again to give the mass

$$m_\Lambda = \frac{\sqrt{E_\Lambda^2 - (p_\Lambda c)^2}}{c^2} = 1116 \text{ MeV}/c^2$$

This is, in fact, how the masses of many unstable subatomic particles are measured. As we discuss in the next section, it is fairly easy to measure the momenta of the decay products as long as they are charged. If the decay masses are known, we can then calculate their energies and hence reconstruct the parent particle's energy and momentum. From these, one can calculate its mass.

2.7 Force in Relativity

We have come a surprisingly long way in relativistic mechanics without defining the notion of force. This reflects correctly the comparative unimportance of forces in relativity. Nonetheless, forces are what change the momentum of a body and we must now discuss them.

Just as with momentum and energy, our first task is to decide on a suitable definition of force. In classical mechanics two equivalent definitions of the force acting on a body are

$$\mathbf{F} = m\mathbf{a} \quad (2.41)$$

and

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \quad (2.42)$$

These are equivalent since $\mathbf{p} = m\mathbf{u}$, with m constant, so $d\mathbf{p}/dt = m\mathbf{a}$. In relativity we have defined \mathbf{p} as $\gamma m\mathbf{u}$, and it is no longer true that $d\mathbf{p}/dt = m\mathbf{a}$. Therefore, we certainly cannot carry over *both* definitions, (2.41) and (2.42), into relativity. In fact, for most purposes the convenient definition of force in relativity is the second of the classical definitions, (2.42).

The total force \mathbf{F} acting on a body with momentum \mathbf{p} is defined as

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \quad (2.43)$$

Clearly, the definition (2.43) reduces to the nonrelativistic definition if the body concerned is moving nonrelativistically. Further, with this definition of \mathbf{F} , the work-energy theorem carries over into relativity, as we prove in the following example.

Example 2.9

Prove that if a mass m , acted on by a total force \mathbf{F} , moves a small distance $d\mathbf{r}$, the change in its energy, dE , equals the work done by \mathbf{F} .

$$dE = \mathbf{F} \cdot d\mathbf{r} \quad (2.44)$$

To prove this, we replace \mathbf{F} by $d\mathbf{p}/dt$ and $d\mathbf{r}$ by $\mathbf{u} dt$ on the right to give

$$\mathbf{F} \cdot d\mathbf{r} = \frac{d\mathbf{p}}{dt} \cdot \mathbf{u} dt \quad (2.45)$$

Now, we will prove in a moment that

$$\frac{d\mathbf{p}}{dt} \cdot \mathbf{u} = \frac{dE}{dt} \quad (2.46)$$

Thus from (2.45) it follows that

$$\mathbf{F} \cdot d\mathbf{r} = \frac{dE}{dt} dt = dE$$

which is the work-energy theorem. It remains only to prove the identity (2.46), as follows: From the “Pythagorean relation,” we know that

$$E = (p^2 c^2 + m^2 c^4)^{1/2}$$

Using the chain rule to differentiate this, we find (Problem 2.36)

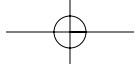
$$\frac{dE}{dt} = \frac{1}{2} (p^2 c^2 + m^2 c^4)^{-1/2} 2\mathbf{p}c^2 \cdot \frac{d\mathbf{p}}{dt} = \frac{\mathbf{p}c^2}{E} \cdot \frac{d\mathbf{p}}{dt} = \mathbf{u} \cdot \frac{d\mathbf{p}}{dt} \quad (2.47)$$

which is (2.46).

The most important force in most applications of relativistic mechanics is the electromagnetic force. It is found experimentally that when a charge q is placed in electric and magnetic fields \mathbf{E} and \mathbf{B} , its relativistic momentum changes at a rate equal to $q(\mathbf{E} + \mathbf{u} \times \mathbf{B})$. Thus, having defined \mathbf{F} as $d\mathbf{p}/dt$, we find that the electromagnetic force is given by the classical formula, often called the **Lorentz force**,

$$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B})$$

Even when the force \mathbf{F} is known, the equation $d\mathbf{p}/dt = \mathbf{F}$ is usually hard to solve for a body's position as a function of time. One important case where it is easily solved is for a charged body in a uniform magnetic field \mathbf{B} . In this case the force $\mathbf{F} = q(\mathbf{u} \times \mathbf{B})$ is perpendicular to \mathbf{u} , and the body's energy is



66 Chapter 2 • Relativistic Mechanics

therefore constant. [This follows from the work-energy theorem (2.44) because \mathbf{F} is perpendicular to $d\mathbf{r}$ as the body moves along.] Therefore, the velocity is constant in magnitude and changes only its direction. In particular, if the body is moving in a plane perpendicular to \mathbf{B} , it moves in a circular path, whose radius R can be found as follows: Since γ is constant, the equation

$$\frac{d\mathbf{p}}{dt} = q\mathbf{u} \times \mathbf{B}$$

becomes

$$\gamma m \frac{d\mathbf{u}}{dt} = q\mathbf{u} \times \mathbf{B}$$

Since $d\mathbf{u}/dt = \mathbf{a}$ and \mathbf{u} is perpendicular to \mathbf{B} , this implies that

$$\gamma ma = quB$$

For motion in a circle, we know that a is the centripetal acceleration u^2/R . (This purely kinematic result is true in relativity for exactly the same reasons as in classical mechanics.) Therefore,

$$\gamma m \frac{u^2}{R} = quB$$

whence, since $\gamma mu = p$,

$$R = \frac{p}{qB} \quad (2.48)$$

This result provides a convenient way to measure the momentum of a particle of known charge q . If we send the particle into a known magnetic field B , then by measuring the radius R of its curved path we can find its momentum p from (2.48).

Example 2.10

A proton of unknown momentum p is sent through a uniform magnetic field $B = 1.0$ tesla (T), perpendicular to \mathbf{p} , and is found to move in a circle of radius $R = 1.4$ m. What are the proton's momentum in MeV/c and its energy in MeV?

The proton's charge is known to be $q = e = 1.6 \times 10^{-19}$ coulomb (C). Thus from (2.48) its momentum is

$$\begin{aligned} p &= qBR = (1.6 \times 10^{-19} \text{ C}) \times (1.0 \text{ T}) \times (1.4 \text{ m}) \\ &= 2.24 \times 10^{-19} \text{ kg} \cdot \text{m/s} \end{aligned}$$

From the list of conversion factors inside the front cover, we find $1 \text{ MeV}/c = 5.34 \times 10^{-22} \text{ kg} \cdot \text{m/s}$. [This was derived in (2.26).] Therefore,

$$p = (2.24 \times 10^{-19} \text{ kg} \cdot \text{m/s}) \times \frac{1 \text{ MeV}/c}{5.34 \times 10^{-22} \text{ kg} \cdot \text{m/s}} = 420 \text{ MeV}/c$$

Since the proton's rest mass is known to be $m = 938 \text{ MeV}/c^2$, we can find its energy from the "Pythagorean relation" (2.38) to be

$$E = \sqrt{(pc)^2 + (mc^2)^2} = 1030 \text{ MeV}$$

2.8 Massless Particles

In this section we consider a question that will probably strike you as peculiar if you have never met it before: In the framework of relativistic mechanics, is it possible to have particles of zero mass, $m = 0$?

In classical mechanics the answer to this question is undoubtedly "no." The classical momentum and kinetic energy of a particle are $m\mathbf{u}$ and $\frac{1}{2}m\mathbf{u}^2$. If $m = 0$, both of these are also zero, and a particle whose momentum and kinetic energy are always zero is presumably nothing at all.

In relativity the answer is not as clear cut. Our definitions of energy and momentum were

$$\mathbf{p} = \gamma m\mathbf{u} \quad (2.49)$$

and

$$E = \gamma mc^2 \quad (2.50)$$

and from these we derived the two important relations

$$E^2 = (pc)^2 + (mc^2)^2 \quad (2.51)$$

and

$$\beta = \frac{u}{c} = \frac{pc}{E} \quad (2.52)$$

Let us consider the last two relations first. If there were a particle with $m = 0$ [and if the relations (2.51) and (2.52) applied to this particle], then (2.51) would imply that

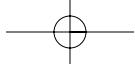
$$E = pc \quad (\text{if } m = 0) \quad (2.53)$$

and (2.52) would then tell us that

$$\beta = 1 \quad \text{or} \quad u = c \quad (\text{if } m = 0)$$

That is, the massless particle would always have speed $u = c$. The converse of this statement is also true. If we discovered a particle that traveled with speed c , by (2.52) it would satisfy $pc = E$, and (2.51) would then require that $m = 0$.

If we turn to our original definitions of \mathbf{p} and E , (2.49) and (2.50), a superficial glance suggests the same result as in classical mechanics. With $m = 0$, (2.49) and (2.50) appear at first to imply that \mathbf{p} and E are zero. However, a closer look shows that this is not so. We have already seen that a massless particle would have to travel at speed c ; and if $u = c$, then $\gamma = \infty$. Thus both



68 Chapter 2 • Relativistic Mechanics

(2.49) and (2.50) have the form $\infty \times 0$, which is indeterminate. Thus our original definitions fail to define \mathbf{p} and E if $m = 0$, but they do not actually contradict the possible existence of particles with $m = 0$.

Apparently, it is logically possible in relativity to have particles with $m = 0$; and, in fact, experiment shows that such particles do exist, the most important example being the photon, or particle of light. In classical physics, light (like all other forms of electromagnetic radiation) was assumed to be a wave in which the energy and momentum were distributed continuously through space. As we will describe in Chapter 4, it was discovered soon after 1900 that the energy and momentum in an electromagnetic wave are actually confined to many tiny localized bundles. These tiny bundles, which have come to be called **photons**, display many of the properties of ordinary particles like the electron; in particular, they have energy and momentum. But unlike the more familiar particles, they travel at the speed of light and, by what we have already said, must therefore have $m = 0$.

Since the definitions (2.49) and (2.50) for momentum and energy cannot be used for massless particles, the question naturally arises as to how we can define and measure these quantities. One simple answer is to consider a process that involves only one massless particle. For example, light shining on an atom can tear loose one of the atom's electrons, a process called photoionization. According to the photon theory of light, photoionization occurs when one of the photons that make up the light collides with the atom and ejects one of the atom's electrons. In the process the photon is absorbed and disappears. For example, the photoionization of the simplest atom, hydrogen (made up of one electron and a proton), can be represented as



where γ is the traditional symbol for a photon and H , e , and p stand for the hydrogen atom, the electron, and the proton. The important point is that the three bodies H , e , and p all have mass and hence have well-defined momenta and energies. Thus if we assume that momentum and energy are conserved in the process (2.54), we can calculate the momentum and energy of the photon from the known values for the other three particles.

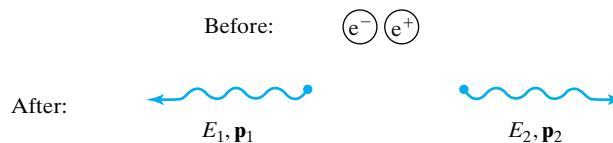
If we compute \mathbf{p} and E for the photon in the manner just described, we find two essential properties. First, the definition is consistent: One obtains the same values for \mathbf{p} and E when identical photons are observed in two or more different processes. Second, the values obtained satisfy (2.53), $pc = E$, which we saw was an essential characteristic of a particle with $m = 0$ and $u = c$.

We will review much more of the evidence for the existence of photons and their zero rest mass in Chapter 4. Today it is almost universally accepted that the photon is a particle that carries energy and momentum and can be treated much like other particles, except that it always travels at speed c and has mass exactly equal to zero.

Until recently, it was generally believed that a second particle, the neutrino, also has zero rest mass. This is a particle observed in the decay of many radioactive nuclei, as we will describe in Chapter 17. However, there is now evidence that the neutrino's mass, although very small (possibly of order 10^{-5} times the electron mass), is not exactly zero. There are theoretical reasons for believing that there is a massless particle called the graviton, which is related to gravity in the same way that the photon is related to light; but there is, as yet, no experimental evidence for the graviton.

Example 2.11

As we will discuss in Chapters 17 and 18, there is a subatomic particle called the *positron*, or *antielectron*, with exactly the same mass as the electron ($0.511 \text{ MeV}/c^2$) but the opposite charge. The most remarkable property of the positron is that when it collides with an electron, the two particles can annihilate one another, converting themselves into two or more photons. Consider the case that the electron and positron are both at rest and that just two photons are produced, with energies E_1 and E_2 and momenta \mathbf{p}_1 and \mathbf{p}_2 (Fig. 2.6). Use conservation of energy and momentum to find the energies, E_1 and E_2 of the two photons.

**FIGURE 2.6**

The annihilation of an electron-positron pair into two photons.

Before the annihilation, the total momentum is zero and the total energy is $2mc^2$, where m is the rest mass of the electron. By conservation of momentum and energy, it follows that

$$\mathbf{p}_1 + \mathbf{p}_2 = 0 \quad (2.55)$$

and

$$E_1 + E_2 = 2mc^2 \quad (2.56)$$

From (2.55) we see that the photons have equal and opposite momenta ($\mathbf{p}_1 = -\mathbf{p}_2$, as was suggested in Fig. 2.6). This means that the photons' energies are equal, $E_1 = E_2$ (since $E_1 = p_1c$ and $E_2 = p_2c$). Thus from (2.56)

$$E_1 = E_2 = mc^2 = 0.511 \text{ MeV}$$

Each photon carries away exactly the rest energy of one electron.

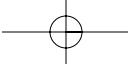
The most remarkable thing about this process is this: In the initial state the total energy is the rest energy of the two particles, $2mc^2$; that is, all the energy is mass energy. In the final state the two photons have no rest mass and there is therefore no mass energy. Thus the process involves 100% conversion of mass energy into another form of energy (electromagnetic, in this case). The observation of such processes is triumphant justification for the claim that the rest energy mc^2 of a mass m is a real energy, not just the result of a whimsical choice for the zero of energy.

This process of electron-positron annihilation with production of two 0.511-MeV photons is observed routinely in particle physics laboratories using artificially produced positrons. It also occurs naturally in the earth's atmosphere where the positrons are produced by cosmic radiation. (The discovery of the positron by the American physicist Carl Anderson involved the observation of such cosmic-ray positrons.) Recently, astronomers have observed photons with exactly 0.511 MeV coming from the center of our galaxy; the conclusion is almost inescapable that somewhere in the galaxy's center positrons are being created and then annihilating with electrons to produce these photons.

Carl Anderson
(1905–1991, American)



Anderson got his PhD from Cal-Tech and spent his entire career as a professor there. His two most important discoveries came from his studies of cosmic rays, the elementary particles, such as protons, that enter the earth's atmosphere from outer space and create a variety of "secondary" particles when they collide with atoms in the atmosphere. Among these secondary cosmic rays, Anderson discovered the positron in 1932 and the muon (described in Chapter 18) in 1935. He won the 1936 Nobel Prize in physics for his discovery of the positron.



70 Chapter 2 • Relativistic Mechanics

A medical application of electron-positron annihilation is the technique called PET or Positron Emission Tomography, in which a patient is injected with a solution containing a positron-emitting radioactive element (for example carbon 11). A suitably chosen solution is attracted to an area of concern (for example, a glucose solution may tend to collect in active areas of the brain), which then emit positrons. The positrons quickly annihilate with nearby electrons, and the outgoing photons can be monitored by a ring of detectors. In this way it is possible to make an accurate map of the areas of interest in the patient.

2.9 When Is Nonrelativistic Mechanics Good Enough?

In most problems — although certainly not all — the equations of classical mechanics are easier to use than their relativistic counterparts. It is therefore important to be able to recognize situations where nonrelativistic mechanics is a good enough approximation and the complications of relativity can be ignored. Loosely speaking, the rule is clear: If a body has speed much less than c at all times, it can be treated nonrelativistically. Unfortunately, this gives us no guidance as to how small the speed must be before it can be considered “much less than c ,” and in fact there is no clear-cut answer. The speed at which relativistic effects must be considered depends on the desired accuracy.

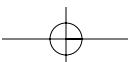
We can illustrate these ideas with two simple examples: First, we saw in Section 1.8 (Example 1.1) that in an airplane traveling at 300 m/s for an hour, relativistic time dilation can affect the plane’s clock by a few nanoseconds ($1 \text{ ns} = 10^{-9} \text{ s}$). In almost any situation this effect is utterly insignificant, and if this is the case, we can ignore relativity. Nevertheless, the Global Positioning System (GPS) described in Section 2.11 requires timings that are accurate to about a nanosecond, and the people who designed this system could certainly not neglect relativistic effects.

As a second example, suppose that we want to know the kinetic energy K of a mass m with speed u . The correct relativistic answer is $K_{\text{rel}} = (\gamma - 1)mc^2$, while its nonrelativistic approximation is $K_{\text{NR}} = \frac{1}{2}mu^2$. We have tabulated both of these, with their percent discrepancy, for various speeds, in Table 2.4. We see that at a speed $u = 0.01c$ (some 300 times the speed of an orbiting satellite) K_{NR} is within 0.01% of the correct K_{rel} . Thus for most purposes, the nonrelativistic answer would be quite good enough when $u = 0.01c$; nevertheless, if for some reason we needed an accuracy better than 0.01%, we would have to work relativistically. At $u = 0.1c$, K_{NR} is still within 1% of the correct K_{rel} , which for some purposes would still be acceptable. At $u = 0.5c$, K_{NR} differs from K_{rel} by 20%, which in most cases would not be acceptable (although, even here, K_{NR} still has the right order of magnitude).

TABLE 2.4

The relativistic and nonrelativistic kinetic energy of a mass m at various speeds u , in units of mc^2 .

u:	0.01 c	0.1 c	0.5 c
$K_{\text{rel}} = (\gamma - 1)mc^2$:	5.004×10^{-5}	5.038×10^{-3}	0.155
$K_{\text{NR}} = \frac{1}{2}mu^2$:	5.000×10^{-5}	5.000×10^{-3}	0.125
% difference:	0.01%	1%	20%



These examples show clearly that whether nonrelativistic mechanics gives an acceptable approximation depends on what one wishes to calculate and with what accuracy. Nevertheless, as a very rough guide for the user of this book, we can state the following three rules:

1. If a particle has speed much less than $0.1c$, it is usually satisfactory to apply nonrelativistic mechanics. For example, we will find that the electron in a hydrogen atom has $u \approx 0.01c$ and we will get excellent results for its energy treating it nonrelativistically.
2. If a particle has a speed of $0.1c$ or more, then, except in very approximate calculations, one should usually use relativistic mechanics.
3. Since they always travel at speed c , massless particles can never be treated nonrelativistically. For example, in considering the photons of light emitted by an atom, we can frequently treat the atom nonrelativistically, but the photons themselves must always be treated relativistically.

Finally, since we often know a particle's energy, rather than its speed, it is convenient to rephrase these rough rules in terms of energy. To decide whether we must use relativity, we compare the particle's kinetic energy K with its rest energy mc^2 . We see from Table 2.4 that if $u \approx 0.1c$, then (very roughly) $K \approx 0.01mc^2$. Thus the first two rules can be rephrased as follows:

1. If a particle's kinetic energy is much less than 1% of its rest energy, it is usually satisfactory to use nonrelativistic mechanics.
2. If the particle has kinetic energy equal to 1% of its rest energy or more, one should usually use relativistic mechanics.

2.10 General Relativity*

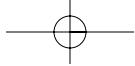
* As we describe here, general relativity is an important part of modern physics. Nevertheless, we will not be using the ideas of this section again, and, if you are pressed for time, you could omit this section without loss of continuity.

Einstein's general relativity, completed in 1915, is the extension of special relativity to include the effects of gravity. We will see that the examination of gravity led Einstein to consider noninertial reference frames. Thus general relativity is more general than special relativity both because it includes gravity *and* because it focuses on noninertial, as well as inertial, reference frames.* Because it is mathematically more complicated than special relativity and because we will have no occasion to use it again in this book, we content ourselves here with a brief qualitative introduction to general relativity.

Inertial Forces

We can learn much about general relativity — just as we did about special relativity — by first examining certain questions within classical physics. In particular, let us consider two classical frames, an inertial frame S and a *noninertial* frame S' , accelerating relative to S with acceleration \mathbf{A} . In the inertial frame S , the equation of motion for a mass m is Newton's second law, $\mathbf{ma} = \sum \mathbf{F}$, where $\sum \mathbf{F}$ is the sum of all forces on the mass. To find the corresponding equation in the noninertial frame S' , we recall the classical

* Experts may object that even special relativity can handle noninertial frames; nevertheless, its primary focus is inertial frames. We should also mention that when we speak of inertial frames here, we mean the inertial frames of special relativity (or of Newtonian mechanics when our discussion is classical).



72 Chapter 2 • Relativistic Mechanics

velocity addition formula, $\mathbf{u}' = \mathbf{u} - \mathbf{v}$. Differentiating, we find that the mass's acceleration as measured in S' is

$$\mathbf{a}' = \mathbf{a} - \mathbf{A} \quad (2.57)$$

since $d\mathbf{v}/dt = \mathbf{A}$, the acceleration of the frame S' . Multiplying (2.57) by m , and putting $m\mathbf{a} = \sum \mathbf{F}$, we find that

$$m\mathbf{a}' = \sum \mathbf{F} - m\mathbf{A}$$

This equation has exactly the form of Newton's second law *except* that in addition to all the forces identified in S , there is an extra force term equal to $-m\mathbf{A}$. Thus Newton's second law is also valid in the noninertial frame S' , *provided* that we recognize that in a noninertial frame every mass must experience an additional “**inertial force**”

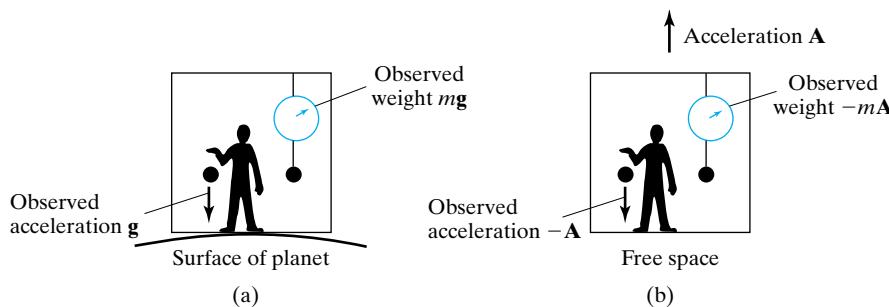
$$\mathbf{F}_{\text{in}} = -m\mathbf{A} \quad (2.58)$$

This inertial force experienced in noninertial frames is familiar in several everyday situations: If we sit in an aircraft accelerating rapidly toward take-off, then from our point of view, there is a force that pushes us back into our seat. If we are standing in a bus that brakes suddenly (\mathbf{A} backward), the inertial force $-m\mathbf{A}$ is forward and can make us fall on our faces if we aren't properly braced. As a car goes rapidly around a sharp curve, the inertial force experienced by its occupants is the so-called centrifugal force that pushes them outward. One can take the view that the inertial force is a “fictitious” force, introduced simply to preserve the form of Newton's second law in noninertial frames. Nevertheless, for an observer in an accelerating frame, it is entirely real.

The Equivalence Principle

The starting point of general relativity is called the equivalence principle and arises from the following observation: The inertial force $\mathbf{F}_{\text{in}} = -m\mathbf{A}$ in (2.58) is proportional to the mass m of the object under consideration, and the same is true of the gravitational force $\mathbf{F}_{\text{gr}} = m\mathbf{g}$ on any mass. That F_{gr} is proportional to m follows from experiments in which two different masses are dropped simultaneously and fall at the same rate — as supposedly tested by Galileo off the leaning tower of Pisa. All objects fall at the same rate only if F_{gr} is proportional to m , since only then does m cancel out of the equation of motion. The proportionality of F_{gr} and m has now been verified to about 1 part in 10^{11} .

That both inertial and gravitational forces are proportional to mass causes a remarkable ambiguity. Suppose that we are in an enclosed cabin, at rest on the surface of a planet with gravitational acceleration \mathbf{g} . Then one of the forces on any mass m is the gravitational attraction of the planet, $m\mathbf{g}$. We could try to verify this, for example, by dropping the mass or by weighing it [Fig. 2.7(a)]. However, a little reflection should convince you that there is no mechanical experiment that can unambiguously confirm the presence of the gravitational force $m\mathbf{g}$. The trouble is that either of the experiments suggested in Fig. 2.7(a) (or any other mechanical experiment) can equally be explained by assuming that our cabin is in gravity-free space but is accelerating upward with $\mathbf{A} = -\mathbf{g}$, as in Fig. 2.7(b). Any effect that we attributed to the

**FIGURE 2.7**

(a) Two experiments designed to verify the existence of the gravitational force $m\mathbf{g}$ on the surface of a planet. **(b)** The same experiments would produce the same results if we were not near any gravitating body, but were instead accelerating with an acceleration $\mathbf{A} = -\mathbf{g}$.

gravitational force $m\mathbf{g}$ can equally be explained by the inertial force $-m\mathbf{A}$. The impossibility of distinguishing between a gravitational force $m\mathbf{g}$ and the equivalent inertial force $-m\mathbf{A}$ (with \mathbf{A} equal but opposite to \mathbf{g}) is called the **principle of equivalence**. In classical physics the equivalence principle applies only to mechanical experiments, but Einstein's general relativity starts from the assumption that the equivalence principle applies to *all* the laws of physics, mechanical and otherwise:^{*}

EINSTEIN'S EQUIVALENCE PRINCIPLE

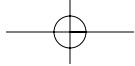
No experiment, mechanical or otherwise, can distinguish between a uniform gravitational field (\mathbf{g}) and the equivalent uniform acceleration ($\mathbf{A} = -\mathbf{g}$).

The general theory of relativity is built on this postulate in much the same way that special relativity was built on its two postulates (Section 1.6). As we will describe briefly, the experimental evidence is such as to convince most physicists that general relativity and the equivalence principle on which it is based are correct.

As one would expect, general relativity agrees with the Newtonian theory of gravity under those conditions where the latter was already known to work well. Specifically, as the gravitating masses get smaller, the differences between Einstein's and Newton's theories approach zero. In fact, even with a mass as large as the sun's, the difference is usually very small. General relativity is of practical importance only for systems that include very large dense masses and in situations requiring high precision, where very small differences may be important.

One branch of physics where general relativity has always been important is cosmology, the study of the structure and evolution of the whole universe. Here the effects of gravity are paramount and *have* to be treated properly (that is, using general relativity). In the last four decades there have been several developments relevant to general relativity, and there has been a burgeoning of interest in the subject. One such development is the discovery of the probable existence of black holes, whose behavior can be analyzed only with the help of general relativity. Another is the success of several experiments with sufficient precision to distinguish between Newton's and Einstein's

*Notice the striking parallel between the general and special theories. In both cases Einstein took a principle that applied to classical mechanics and made the bold assumption that it applied to all physical laws.



74 Chapter 2 • Relativistic Mechanics

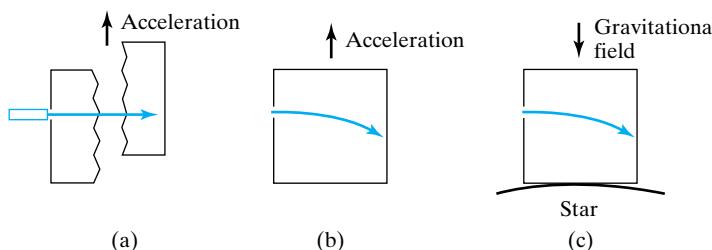
(and other) theories of gravity. (And, incidentally, all such experiments favor Einstein's theory.) Finally, navigational systems using satellites, such as the Global Positioning System of the U.S. Air Force, have become so precise that tiny corrections for the effect of gravity on time have to be made using general relativity.

A remarkable feature of general relativity is that the effects of gravity are incorporated into the *geometry* of space. Thus, instead of saying that the sun exerts forces on the planets causing them to follow their curved orbits, we say that the gravitational field of the sun causes a curvature of space, and it is this curvature that is responsible for the curved orbits of the planets. These ideas are expressed mathematically in the language of *differential geometry*, but to describe how this works would take us too far afield. Instead, we will just describe a few of the theory's simpler consequences.

The Bending of Light by Gravity

We can use the equivalence principle to predict the effects of a gravitational field in several experiments. Suppose, for example, that we shine a beam of light horizontally across the gravitational field of a massive star. The equivalence principle guarantees that the outcome of this experiment must be the same as if we were to observe the beam of light from the equivalent accelerated frame in gravity-free space. Accordingly, we consider first a flash of light that is shone through the window of a cabin S' that accelerates upward relative to an inertial frame S , in free space, as in Fig. 2.8. In part (a) we see the experiment as viewed in the inertial frame S , where the light travels in a horizontal path, but the cabin moves upward with increasing speed; evidently, the light hits the far wall at a point lower than the point at which it entered. Part (b) shows the same experiment as seen inside the accelerating cabin S' ; in this frame the path of the light is curved downward. Finally, the equivalence principle guarantees that the behavior observed in the upwardly accelerating frame of part (b) must be the same as would be observed in the downward gravitational field of part (c). We conclude that light traveling across the gravitational field of a massive body must be deflected downward, toward the body.

Using general relativity, Einstein calculated that a light ray skimming past the surface of the sun should be deflected by 1.75 arc-seconds.* This means that the apparent position of a star closely aligned with the sun's rim should be shifted by 1.75 arc-seconds from its true position, as shown in



*This calculation is actually quite complicated: The simple argument based on the equivalence principle gives only half of the deflection, and the other half comes from a subtle geometric effect of general relativity.

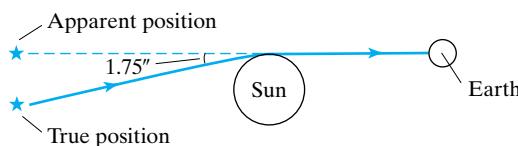


Fig. 2.9. This shift is very hard to observe, since a star closely aligned with the sun is invisible because of the sun's own brighter light. However, it was observed during the solar eclipse of 1919, and although the measurement was only about 30% accurate, it provided early experimental support for general relativity. In the early 1970s the deflection was measured much more accurately using radio telescopes, and the results agreed with general relativity within about 1%. More recently, experiments using radar reflected from planets and spacecraft have provided results that agree within about 0.1%.

The Gravitational Redshift

A second effect of gravity on light is the gravitational redshift. Imagine a beam of light fired upward from the floor to the ceiling of a cabin in the gravitational field of the earth (or other massive body), as in Fig. 2.10(a). The question we consider is this: If the frequency of the source on the floor is f_0 , what is the frequency of the light received by a detector on the ceiling? To answer this question, we have only to imagine the cabin to be in gravity-free space, with acceleration $\mathbf{A} = -\mathbf{g}$, as in Fig. 2.10(b). We consider two inertial frames, S , in which the source is at rest when the light is emitted, and S' , the frame in which the detector is at rest when the light is received.* The emitted frequency f_0 is the frequency measured in S , and the received frequency f is that in S' . During the time that the light travels from source to detector, the whole cabin is accelerating upward. This means that frame S' is moving upward relative to S , and the frequency f measured in S' is therefore redshifted compared to the frequency f_0 measured in S . (This is the Doppler effect, described in Section 1.14.) By the equivalence principle, the same conclusion applies in the gravitational field of Fig. 2.10(a). Therefore, light is redshifted as it travels upward in a gravitational field. By a similar argument, it is blueshifted if it travels downward. The shifts concerned are usually very small, and this prediction was not accurately confirmed until 1960, when R. V. Pound and G. A. Rebka verified it using high-frequency radiation that traveled up and down a tower at Harvard. It has subsequently been checked with even greater precision (about 0.02%) using radiation sent from high-altitude rockets down to the earth's surface.

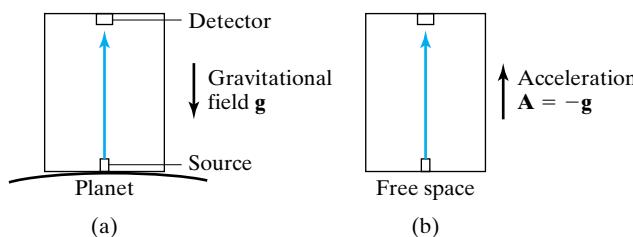


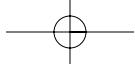
FIGURE 2.9

Light from the star on the left is deflected by 1.75 arc-seconds as it skims past the sun. The star's apparent position is therefore 1.75 arc-seconds above its true position.

FIGURE 2.10

(a) In the gravitational redshift experiment, light is shone upward in a gravitational field. (b) The experiment of (a) is equivalent to this experiment, in which the cabin is accelerating upward in free space. The rest frame of the detector (at the time of detection) moves upward relative to the rest frame of the source (at the time of emission). Therefore, the light received is redshifted.

*Note that neither of these inertial frames is the frame of the accelerating elevator (which is, of course, noninertial). S coincides with the elevator at the moment of emission; S' coincides with the elevator at the moment of reception of the signal.



76 Chapter 2 • Relativistic Mechanics

Example 2.12

It is not hard to show that the fractional frequency shift of light traveling vertically down through a height h in a gravitational field g is $\Delta f/f = gh/c^2$. (See Problem 2.45.) Given that the tower used by Pound and Rebka was about 20 meters tall, what fractional frequency did they expect to see?

The expected shift is $\Delta f/f = gh/c^2 \approx (10 \text{ m/s}^2) \times (20 \text{ m})/(3 \times 10^8)^2 \approx 2 \times 10^{-15}$. Pound and Rebka used photons emitted by the nuclei of radioactive cobalt 57. They were able to measure this astonishingly tiny shift in the photons' frequency by exploiting the Mössbauer effect, in which the nuclei that emit (and later absorb) the photons are held fixed in a crystalline lattice to reduce the Doppler broadening caused by the nuclear recoil.

The Precession of Mercury's Orbit

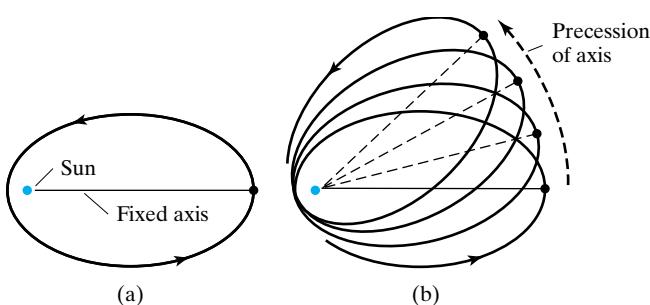
The first test of general relativity was based on the discovery made in 1859 that the planet Mercury does not move in the perfect, fixed ellipse predicted by Newtonian theory [Fig. 2.11(a)]; instead, the orbit precesses in such a way that its axis rotates slowly (by 43 arc-seconds per century) as in Fig. 2.11(b).^{*} This anomalous motion had suggested to some astronomers that Mercury's orbit was being disturbed by a hitherto undetected planet, which was even given the name Vulcan. However, careful searches failed to find Vulcan. Some 50 years later, when he completed his general theory of relativity, Einstein found that the theory predicted a slow precession of all planetary orbits. For Mercury, the theory predicted a precession in almost perfect agreement with the (present) observed 42.98 arc-seconds per century. For the other planets, the predicted values were much smaller, but several of these have now been measured and agree well with the predictions.

Black Holes

The prediction of general relativity that has most caught the public imagination is the black hole. This is a very heavy and dense star, whose gravitational field is so strong that no light — or anything else — can escape from its interior. A black hole is formed when a star of several solar masses exhausts its

FIGURE 2.11

- (a) According to Newtonian theory, a planet should move round an ellipse whose axis is fixed.
- (b) The axis of the ellipse actually rotates slowly, in agreement with general relativity. The effect is greatly exaggerated in this picture; the actual precession is greatest for Mercury, whose axis advances by just 43 arc-seconds per century.



*The situation is actually more complicated: The other planets distort Mercury's orbit much more than this tiny effect, but when all these well-understood distortions were subtracted out, there remained an unexplained advance of 43 arc-seconds per century. Because an orbit's axis is conveniently specified by its perihelion (the point of closest approach to the sun), this effect is usually referred to as the advance of Mercury's perihelion.

supply of nuclear fuel. Without the fuel to maintain the pressure needed to support it, the star begins to collapse and continues to do so until its radius approaches a value called the Schwarzschild radius, R_S . As $r \rightarrow R_S$, the rate of collapse slows down (as observed from far away), the redshift of light from the star grows indefinitely, and light from inside R can no longer escape at all. (For an estimate of the Schwarzschild radius, see Problem 2.46.) Since light from inside a black hole cannot escape, direct evidence for black holes is hard to come by. Nevertheless, there is now strong circumstantial evidence for their existence. For example, if an ordinary star were caught in orbit around a black hole, matter would be torn from the star by the black hole. Before it disappears inside the black hole, this falling matter should reach such high energies that its collisions should produce X-rays. Certain astronomical X-ray sources have been observed that fit the predictions of this model well. In addition, it is thought that extremely massive black holes may be responsible for the intense radiation from the superenergetic and distant objects called quasars. Black holes are discussed further in Section 13.10.

Gravitational Waves

Another prediction of general relativity that has attracted much attention in the last few decades is the possibility of gravitational waves. Unlike the Newtonian theory, general relativity predicts that accelerating masses should radiate gravitational waves, just as accelerating electric charges radiate electromagnetic waves (like the radio waves from oscillating charges in a radio antenna). However, even the most violent cosmic events produce very feeble gravitational waves, and although there have been several heroic attempts at detection, there have so far been no reproducible, direct observations of gravitational waves. On the other hand, there is strong indirect evidence, in that certain rapidly rotating star systems have been observed to be losing energy at precisely the rate that should result from their gravitational radiation. Several experiments are under construction to detect gravitational waves. One such is called LIGO (for Laser Interferometer Gravity-wave Observatory) and consists of a Michelson interferometer with arms 4 kilometers long. Any passing gravitational wave should change the length of one of the arms enough to cause an observable shift in the interference pattern. Two of these are under construction, one in Washington State and the other in Louisiana. Scientists hope that such interferometers will not only be able to detect gravitational waves but also give us an entirely new kind of “telescope” for studying the universe.

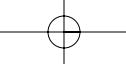
Overview

The observation in 1919 of the gravitational deflection of light passing the sun has been described as the first scientific media event. It drew public attention to relativity theory and helped make Einstein the best known scientific celebrity of all time. Nevertheless, for the next 40 years general relativity attracted less and less attention, and it remained a specialized and even obscure branch of physics. Then, sometime in the 1950s, it entered a renaissance, with the new generation of experiments that were sensitive enough to test it, and with the discovery of phenomena, like black holes, that absolutely require general relativity for their understanding. Today, it is, once again, undeniably an important part of modern physics.

Stephen Hawking
(born 1942, English)



Definitely the best known of the leaders of the new interest in general relativity, Hawking has made seminal contributions to the unification of gravity and quantum theory, and to the theory of black holes. While still a student, he was afflicted by amyotrophic lateral sclerosis (Lou Gehrig's disease), but he has continued his research and his popularization of physics in spite of the unimaginable handicaps of immobility and lack of speech.



78 Chapter 2 • Relativistic Mechanics

2.11 The Global Positioning System: An Application of Relativity*

*This section can be omitted without loss of continuity

In our daily lives here on small, slow-moving planet earth, the effects of special and general relativity are almost always too small to notice. But in recent years a new consumer technology has been developed that relies on relativistic corrections for its proper functioning; this is the Global Positioning System, or GPS. With a GPS receiver, which can be purchased at a hardware store for about the cost of a cell phone, you can quickly determine your location on earth (latitude, longitude, and elevation) to within a meter or so. To understand the basic principles of the GPS, you do not need to understand relativity. However, the fantastic accuracy of the system depends crucially on several relativistic effects.

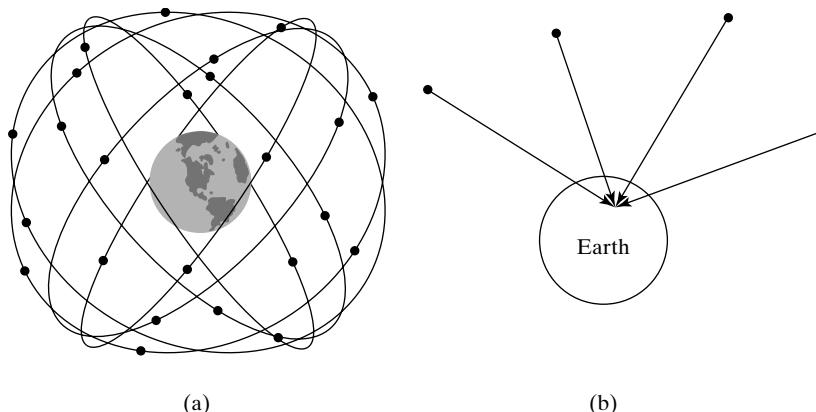
The Global Positioning System consists of 24 satellites, each in a high circular orbit, 20,200 km above the earth's surface. Each of these satellites, developed by the U.S. military and placed in orbit over the period 1977–1994, contains an atomic clock that maintains the exact time within a few nanoseconds. Each satellite continuously broadcasts its precise position and the time (according to its own clock). At least four satellites are visible at any time from any spot on earth, and a handheld GPS receiver listens to the broadcasts of these satellites and computes its position from this information. See Fig. 2.12.

The receiver determines its location by computing the distance to each of the satellites overhead. The radio signal from any one satellite travels to the receiver at the speed of light, c , covering the distance r from satellite to receiver in a time $\Delta t = r/c$. (Δt is about 0.06 seconds for a satellite directly overhead.) The receiver finds Δt as the discrepancy between its own time and the satellite's time signal, and then calculates $r = c \Delta t$. This tells the receiver that it lies somewhere on a sphere of radius r centered on the satellite's known position. In the same way, the receiver establishes that it lies on four spheres of known centers and radii. Two of these spheres intersect in a circle; the third intersects this circle in two points, and the fourth determines the position uniquely.

Our discussion so far is actually an oversimplification because we assumed that the receiver knows its own time exactly, for comparison with the four satellite signals. This would require that the receiver contain an atomic clock costing many tens of thousands of dollars. Instead, the receiver uses the four satellite signals to *calculate* its own time. To see how this works, notice

FIGURE 2.12

The Global Positioning System.
(a) The constellation of 24 GPS satellites moves in 6 orbital planes 20,200 km above the surface of the earth. Each satellite orbits the earth in 12 hours. **(b)** Any GPS receiver on earth can receive signals from at least four satellites at any time. These signals allow the receiver to compute the four unknowns, x, y, z (its position), and the time t .



that there are four unknowns, namely the position $\mathbf{r} = (x, y, z)$ of the receiver and the time t at which the signals are detected. The known positions and times of the four satellites provide four equations:

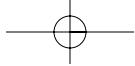
$$|\mathbf{r} - \mathbf{r}_j| = c(t - t_j), \quad j = 1, 2, 3, 4 \quad (2.59)$$

where \mathbf{r}_j and t_j are the positions and times of the four satellites. These four equations can be solved to give the four unknowns, and the receiver can tell us not only our position but also the exact time.

The Global Positioning System has an accuracy of about a meter, but several corrections must be applied to the simple calculation above to achieve this accuracy. Among these are corrections for the index of refraction of the earth's atmosphere, which slows the propagation of the radio signals. Corrections due to relativistic effects are also essential. The simplest correction due to special relativity is already contained in equations (2.59), namely, that the speed of light is a constant, c , independent of the relative motion of the source and observer. The satellites are moving relative to the surface of the earth with a speed of about 6 km/s. We leave it as an exercise (Problem 2.47) to show that use of the (incorrect) classical velocity-addition formula would lead to positioning errors of 100 m or so. A correction for the time dilation of the moving satellite clocks is also necessary (Problem 2.48). The largest correction from general relativity is due to the gravitational redshift, which causes the atomic clocks on the satellites to run slightly faster than clocks on the earth (Problem 2.49). This effect, if unaccounted for, would also lead to positioning errors of a few hundred meters.

CHECKLIST FOR CHAPTER 2

CONCEPT	DETAILS
Rest mass	Mass of object measured when (nearly) at rest (Sec. 2.2)
Relativistic momentum	$\mathbf{p} = \gamma m\mathbf{u}$ (2.4)
Relativistic energy	$E = \gamma mc^2$ (2.9)
Rest energy	$E = mc^2$ (2.12)
Kinetic energy	$K = E - mc^2 = (\gamma - 1)mc^2$ (2.13)
First "useful relation"	$\beta = pc/E$ (2.22)
The "Pythagorean relation"	$E^2 = (pc)^2 + (mc^2)^2$ (2.23)
eV and MeV	Units of energy, 1 eV = 1.6×10^{-19} J, 1 MeV = 10^6 eV
MeV/c and MeV/c^2	Units of momentum ($1 \text{ MeV}/c = 5.34 \times 10^{-22} \text{ kg} \cdot \text{m/s}$) and mass ($1 \text{ MeV}/c^2 = 1.78 \times 10^{-30} \text{ kg}$)
Binding energy, B	$B =$ energy to pull a bound system apart into its separate constituents (2.35)
Force, \mathbf{F}	$\mathbf{F} = d\mathbf{p}/dt$ (2.43)
Work-energy theorem	$dE = \mathbf{F} \cdot d\mathbf{r}$ (2.44)
Lorentz force on a charge q	$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B})$
Circular motion of a charge in a magnetic field	$R = p/qB$ (2.48)
Massless particles	$u = c$ always, and $E = pc$ (2.53)



80 Chapter 2 • Relativistic Mechanics

General relativity*

Inertial forces, equivalence principle, bending of light by gravity, gravitational redshift, precession of mercury's orbit, black holes, gravitational waves (Sec. 2.10)

The Global Positioning System*

Sec. 2.11

PROBLEMS FOR CHAPTER 2

SECTION 2.3 (Relativistic Momentum)

- 2.1** • The mass of an electron is about 9.11×10^{-31} kg. Make a table showing an electron's momentum, both the correct relativistic momentum and the nonrelativistic form, at speeds with $\beta = 0.1, 0.5, 0.9, 0.99$.
- 2.2** • What is the momentum of a 1000-kg space probe when traveling at the speed needed to escape the earth's gravitational attraction (about 10,000 m/s)? What is the percent error from using the nonrelativistic formula $p = mv$ for momentum?
- 2.3** • How fast must a body be traveling if its correct relativistic momentum is 1% greater than the classical $p = mv$?
- 2.4** • Consider the collision between the two billiard balls of Fig. 2.1. Table 2.1 gives the four velocities measured in frame S . (a) Use the classical velocity-addition formula, Eq. (1.41), to verify that the velocities measured in S' would be as given in Table 2.2 if S and S' were related classically. (b) Verify that the correct relativistic answers given in Table 2.3 reduce to the values of Table 2.2 if all speeds are much less than c .
- 2.5** •• Consider the collision between the two billiard balls of Fig. 2.1. The four velocities involved, as measured in S , are given in Table 2.1. Use the velocity transformations (1.43) and (1.44) to verify that the four velocities measured in S' are as given in Table 2.3. Verify that the total classical momentum as measured in S' (that is, $\sum m\mathbf{v}'$) is *not* conserved.
- 2.6** ••• (a) Consider the collision of the two billiard balls of Fig. 2.1. Make a table similar to Table 2.1 but showing the relativistic momenta of the balls before and after the collision, as measured in S . Verify that, as seen in S , the total relativistic momentum is conserved. (b) Now, using the velocities of Table 2.3, make a table showing the four relativistic momenta as measured in S' . Verify that the total relativistic momentum is also conserved in S' . (Be careful! Remember that the factor γ in the definition $p = \gamma mu$ depends on u . Therefore, with four different velocities, there are in general four different factors γ . It takes some courage to wade through the algebra here, but it is rewarding to see the momenta come out equal.)

SECTION 2.4 (Relativistic Energy)

- 2.7** • At what speed would a body's relativistic energy E be twice its rest energy mc^2 ?
- 2.8** • An electron (rest mass 9×10^{-31} kg) is moving at $0.6c$. What is its energy E ? At this speed, what fraction of its energy is rest energy?

- 2.9** • At what speed is a body's kinetic energy equal to twice its rest energy?

- 2.10** • We saw that the relativistic momentum $\mathbf{p} = \gamma m\mathbf{u}$ of a mass m can also be expressed as

$$\mathbf{p} = m \frac{d\mathbf{r}}{dt_0} \quad (2.60)$$

where dt_0 denotes the *proper* time between two neighboring points on the body's path and has the same value for all observers. Show that the relativistic energy $E = \gamma mc^2$ can similarly be rewritten as

$$E = mc^2 \frac{dt}{dt_0} \quad (2.61)$$

The relations (2.60) and (2.61) make it easy to see how \mathbf{p} and E transform from one inertial frame to another. (See Problem 2.14.)

- 2.11** • For the elastic head-on collision of Example 2.3 (Section 2.4), solve the nonrelativistic equations (2.14) and (2.15) and verify the solutions (2.16) and (2.17).

- 2.12** • Consider a relativistic, elastic, head-on collision of two particles as in Example 2.3 (Section 2.4), but suppose that the second particle is much heavier than the first ($m_1 \ll m_2$). Using (2.19), show that $u_3 \approx -u_1$. That is, the first particle bounces back with its speed unchanged (just as in the nonrelativistic case).

- 2.13** •• If one defines a variable mass $m_{\text{var}} = \gamma m$, the relativistic momentum $\gamma m\mathbf{u}$ becomes $m_{\text{var}}\mathbf{u}$, which looks more like the classical definition. Show, however, that the relativistic kinetic energy $(\gamma - 1)mc^2$ is *not* equal to $\frac{1}{2}m_{\text{var}}u^2$.

- 2.14** ••• (a) Suppose that a mass m has momentum \mathbf{p} and energy E , as measured in a frame S . Use the relations (2.60) and (2.61) (Problem 2.10) and the known transformation of $d\mathbf{r}$ and dt to find the values of \mathbf{p}' and E' as measured in a second frame S' traveling with speed v along Ox . (Notice that apart from some factors of c , the quantities \mathbf{p} and E transform just like \mathbf{r} and t . Remember that dt_0 has the same value for all observers.) (b) Use the results of part (a) to prove the following important result: If the total momentum and energy of a system are conserved as measured in one inertial frame S , the same is true in any other inertial frame S' .

- 2.15** ••• Consider the elastic, head-on collision of Example 2.3 (Section 2.4). The algebra leading to Eq. (2.19) for u_3 is surprisingly messy and not very illuminating. To simplify it, consider the case that the two masses are equal. Write down the equations for energy and

Problems for Chapter 2 **81**

momentum conservation, then prove that the final velocities are $u_3 = 0$ and $u_4 = u_1$. (In this special case with equal rest masses, the relativistic result agrees with the nonrelativistic result.)

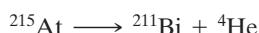
[*HINT:* Remember that the factor γ that appears in the definitions of \mathbf{p} and E depends on velocity; thus, with four different velocities, one has four different values of γ . The identity $\gamma^2(1 - \beta^2) = 1$ will be useful.]

SECTION 2.5 (Two Useful Relations)

- 2.16** • By squaring the definitions (2.20) and (2.21) of \mathbf{p} and E , verify the “Pythagorean relation” $E^2 = (pc)^2 + (mc^2)^2$.
- 2.17** • A nuclear particle has mass $3 \text{ GeV}/c^2$ and momentum $4 \text{ GeV}/c$. **(a)** What is its energy? **(b)** What is its speed? ($1 \text{ GeV} = 10^9 \text{ eV}$.)
- 2.18** • A particle is observed with momentum $500 \text{ MeV}/c$ and energy 1746 MeV . What is its speed? What is its mass (in MeV/c^2 and in kg)?
- 2.19** • A proton (rest mass $938 \text{ MeV}/c^2$) has kinetic energy 500 MeV . What is its momentum (in MeV/c and in $\text{kg} \cdot \text{m/s}$)? How fast is it traveling?
- 2.20** • At the Stanford Linear Accelerator, electrons are accelerated to energies of 50 GeV ($1 \text{ GeV} = 10^9 \text{ eV}$). **(a)** If this energy were classical kinetic energy, what would be the electrons’ speed? (Take the electrons’ mass to be $0.5 \text{ MeV}/c^2$.) **(b)** Calculate γ and hence find the electrons’ actual speed.
- 2.21** • The rest energy of a certain nuclear particle is 5 GeV ($1 \text{ GeV} = 10^9 \text{ eV}$), and its kinetic energy is found to be 8 GeV . What is its momentum (in GeV/c), and what is its speed?
- 2.22** • **(a)** Using the two identities (2.22) and (2.23) prove that the speed of any particle with $m > 0$ is always less than c . **(b)** If there could be particles with $m = 0$ (as in fact there can), prove from the same two identities that such particles would always have speed equal to c .

SECTION 2.6 (Conversion of Mass to Energy)

- 2.23** • When the radioactive nucleus of astatine 215 decays, it tears the whole atom into two atoms, bismuth 211 and helium 4.



(This type of decay is called α decay because the helium 4 nucleus is often called an alpha particle.) The masses of the three atoms are

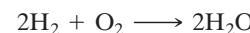
$$\begin{aligned} ^{215}\text{At: } & 3.57019 \times 10^{-25} \text{ kg} \\ ^{211}\text{Bi: } & 3.50358 \times 10^{-25} \text{ kg} \\ ^4\text{He: } & 0.06647 \times 10^{-25} \text{ kg} \end{aligned}$$

What is the kinetic energy released in the decay (in joules and in MeV)?

- 2.24** • If an electron and proton (both initially at rest and far apart) come together to form a hydrogen atom,

13.6 eV of energy is released (mostly as light). By how much does the mass of an H atom differ from the sum of the electron and proton masses? What is the fractional difference $\Delta M/(m_e + m_p)$?

- 2.25** • When two molecules of hydrogen combine with one molecule of oxygen to form two water molecules,

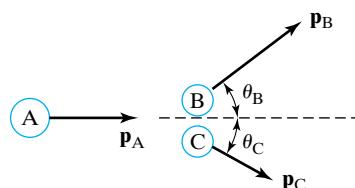


the energy released is 5 eV. **(a)** What is the mass difference between the three original molecules and the two final ones? **(b)** Given that the water molecule has mass about $3 \times 10^{-26} \text{ kg}$, what is the fractional change in mass, $\Delta M/(\text{total mass})$? (Does it matter significantly whether you use the initial or final total mass?) **(c)** If one were to form 10 g of water by this process, what would be the total change in rest mass?

- 2.26** • A subatomic particle A decays into two identical particles B; that is, $A \rightarrow B + B$. The two B particles are observed to have exactly equal and opposite momenta of magnitude p . **(a)** What can you deduce about the velocity of A just before the decay? **(b)** Derive an expression for the mass, m_A , of A in terms of m_B and p .

- 2.27** • A lambda particle (Λ) decays into a proton and a pion, $\Lambda \rightarrow p + \pi$, and it is observed that the proton is left at rest. **(a)** What is the energy of the pion? **(b)** What was the energy of the original Λ ? (The masses involved are $m_\Lambda = 1116$, $m_p = 938$, and $m_\pi = 140$, all in MeV/c^2 . As is almost always the case, your best procedure is to solve the problem algebraically, in terms of the symbols m_Λ , m_p , m_π , and, only at the end, to put in numbers.)

- 2.28** • A particle A moving with momentum $\mathbf{p}_A \neq 0$ decays into two particles B and C as in Fig. 2.13. **(a)** Prove that the three momenta \mathbf{p}_A , \mathbf{p}_B , \mathbf{p}_C lie in a plane. **(b)** If $m_B = m_C$ and if it is found that $\theta_B = \theta_C$, prove that particles B and C must have equal energies. **(c)** If it is found that $\theta_B = 0$, prove that $\theta_C = 0$ or 180° .

**FIGURE 2.13**

(Problems 2.28 and 2.39)

- 2.29** • The K^0 meson is a subatomic particle of rest mass $m_K = 498 \text{ MeV}/c^2$ that decays into two charged pions, $K^0 \rightarrow \pi^+ + \pi^-$. (The π^+ and π^- have opposite charges but exactly the same mass, $m_\pi = 140 \text{ MeV}/c^2$.) A K^0 at rest decays into two pions. Use conservation of energy and momentum, to find the energies, momenta and speeds of the two pions. (Give algebraic answers, in terms of the symbols m_K and m_π first; then put in numbers.)

82 Chapter 2 • Relativistic Mechanics

- 2.30** •• Many problems in relativity are best solved by viewing them first in a cleverly chosen reference frame. Here is an example: A K° meson (see Problem 2.29) traveling at $0.9c$ decays into a π^+ and π^- , sending the π^+ exactly forward and the π^- exactly backward. Using the results of Problem 2.29, find the velocities of the two pions.

[*HINT:* Let S be the frame of this problem and S' the K° rest frame. You can use the results of Problem 2.29 in S' and then use the velocity transformation to find the velocities in S .]

- 2.31** •• A particle of unknown mass M decays into two particles of known masses $m_1 = 0.5 \text{ GeV}/c^2$ and $m_2 = 1 \text{ GeV}/c^2$, whose momenta are measured to be $\mathbf{p}_1 = 2 \text{ GeV}/c$ along the positive y axis and $\mathbf{p}_2 = 1.5 \text{ GeV}/c$ along the positive x axis. ($1 \text{ GeV} = 10^9 \text{ eV}$) Find the unknown mass M and its speed.

- 2.32** •• A mad scientist claims to have observed the decay of a particle of mass M into two identical particles of mass m with $M < 2m$. In response to the objection that this violates conservation of energy, he retorts that if M was traveling fast enough, it could easily have energy greater than $2mc^2$ and hence could decay into the two particles of mass m . Show that he is wrong. (He has forgotten that momentum as well as energy must be conserved. You can analyze this problem in terms of these two conservation laws, but it is much easier to view the proposed reaction from the rest frame of the particle M .)

SECTION 2.7 (Force in Relativity)

- 2.33** • An electron with kinetic energy 1 MeV enters a uniform magnetic field $B = 0.1 \text{ T}$ (perpendicular to the electron's velocity). What is the radius of the resulting circular orbit? [Take the electron's rest mass to be $0.5 \text{ MeV}/c^2$. Don't forget that to use Eq. (2.48), you must express p in the proper SI units, which are $\text{kg} \cdot \text{m}/\text{s}$. The needed conversion factor can be found inside the front cover.]

- 2.34** • An electron (rest mass $0.5 \text{ MeV}/c^2$) traveling at $0.7c$ enters a magnetic field of strength of 0.02 T and moves on a circular path of radius R . **(a)** What would be the value of R according to classical mechanics? **(b)** What is R according to relativity? (The fact that the observed radius agrees with the relativistic answer is good evidence in favor of relativistic mechanics.)

- 2.35** •• Consider the relativistic form of Newton's second law, $\mathbf{F} = d\mathbf{p}/dt$, for a single mass m with velocity \mathbf{u} and momentum $\mathbf{p} = \gamma m\mathbf{u}$. **(a)** Prove that if \mathbf{F} is always perpendicular to \mathbf{u} (as is the case if \mathbf{F} is the magnetic force on a charged particle), then γ is constant and we can write $\mathbf{F} = \gamma m\mathbf{a}$, where \mathbf{a} is the acceleration $d\mathbf{u}/dt$. **(b)** If, instead, \mathbf{F} is parallel to \mathbf{u} , show that $\mathbf{F} = \gamma^3 m\mathbf{a}$. **(c)** Rewrite your results in terms of the variable mass $m_{\text{var}} = \gamma m$. Note that in case (a) the result ($\mathbf{F} = m_{\text{var}}\mathbf{a}$) looks just like the nonrelativistic form of Newton's second law, but in case (b) it does not.

- 2.36** •• In deriving the work-energy theorem (Example 2.9 in Section 2.7), we used the chain rule to find the derivative dE/dt in Eq. (2.47). This is actually a

rather subtle point because E is a function of three variables, $E(p_x, p_y, p_z)$, and the relevant form of the chain rule is

$$dE = \frac{\partial E}{\partial p_x} dp_x + \frac{\partial E}{\partial p_y} dp_y + \frac{\partial E}{\partial p_z} dp_z$$

where the three derivatives are partial derivatives. Use this equation to fill in the details in Eq. (2.47). (If you don't know about partial derivatives, don't try this problem now. We will describe partial differentiation carefully in Chapter 8.)

SECTION 2.8 (Massless Particles)

- 2.37** • The uncharged pion, π° , is a subatomic particle that is closely related to the charged pions discussed previously. The π° has mass $135 \text{ MeV}/c^2$ (very nearly the same as the charged pions, whose mass is $140 \text{ MeV}/c^2$) and decays into two photons, $\pi^{\circ} \rightarrow \gamma + \gamma$. Assuming that the pion was at rest, what is the energy of each photon?

- 2.38** •• A neutral pion traveling along the x axis decays into two photons, one being ejected exactly forward, the other exactly backward. The first photon has three times the energy of the second. Prove that the original pion had speed $0.5c$.

- 2.39** •• Consider the decay $A \rightarrow B + C$ shown in Fig. 2.13 (Problem 2.28). Suppose that both B and C are massless and that $\theta_B = \theta_C$. Use conservation of energy and momentum to prove that $\cos \theta_B = \beta$ where β is the dimensionless "velocity" (u/c) of particle A .

- 2.40** •• The positive pion decays into a muon and a neutrino, $\pi^+ \rightarrow \mu^+ + \nu$. The pion has rest mass $m_{\pi} = 140 \text{ MeV}/c^2$, the muon has $m_{\mu} = 106 \text{ MeV}/c^2$, while the neutrino has $m_{\nu} \approx 0$. (Strictly speaking, there are several kinds of neutrino; the one discussed here is denoted ν_{μ} to indicate that it is the one produced with a muon.) Assuming the original pion was at rest, use conservation of energy and momentum to show that the speed of the muon is given by

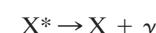
$$\frac{u}{c} = \frac{(m_{\pi}/m_{\mu})^2 - 1}{(m_{\pi}/m_{\mu})^2 + 1}$$

Evaluate this numerically.

- 2.41** ••• An atom (or a nucleus) is ordinarily found in its *ground state*, or state of lowest energy. However, by supplying some energy, one can lift it to an *excited state*. An excited state of an atom X is sometimes denoted X^* and, because of the additional energy ΔE , has a mass m^* slightly greater than that of the ground state (m):

$$m^* = m + \Delta m$$

where $\Delta m = \Delta E/c^2$. If left in isolation, the excited state X^* usually drops back to the ground state emitting a single photon.



If the atom were immovable, the photon would carry off all the additional energy.

$$E_\gamma = \Delta mc^2$$

In reality, conservation of momentum requires the atom to recoil, so that a little of the energy goes to kinetic energy of the recoiling atom. (a) Using conservation of momentum and energy and assuming that the excited atom X^* was at rest, show that

$$E_\gamma = \Delta mc^2 \left(1 - \frac{\Delta m}{2m^*} \right)$$

(b) The energy needed to lift a hydrogen atom from its ground state to the lowest excited state is 10.2 eV. Evaluate the fraction $\Delta m/m^*$ for this state. (Does it make a significant difference whether you use m or m^* in the denominator?) What percentage of the available energy Δmc^2 goes to the photon in the decay of this excited state? (Your result illustrates what a good approximation it usually is to assume that *all* the energy goes to the photon in an atomic transition.)

SECTION 2.9 (When is Nonrelativistic Mechanics Good Enough?)

- 2.42** • If u is much less than c , the binomial approximation ($\gamma = 1 + \beta^2/2$) shows that the relativistic kinetic energy, $K_{\text{rel}} = (\gamma - 1)mc^2$ is about equal to the nonrelativistic $K_{\text{NR}} = \frac{1}{2}mu^2$. An even better approximation is to keep the first *three* terms in the binomial series for γ . (The binomial series can be found in Appendix B.) Show in this way that the difference between K_{rel} and K_{NR} is about $3\beta^4 mc^2/8$. Express this difference as a fraction of K_{NR} and find the maximum value of β for which the difference is less than 1%.

- 2.43** • Consider a relativistic particle of mass m and kinetic energy K . Derive an expression for the particle's speed u in terms of K and m . Make a table showing both the correct value of u and the value you would get if you assumed that $K = \frac{1}{2}mu^2$, for $K = 0.001 mc^2, 0.01 mc^2, 0.1 mc^2$, and mc^2 . Calculate the percent discrepancy for each case. Approximately what is the largest value of K for which a nonrelativistic calculation of u is within 1% of the correct value?

- 2.44** • The cyclotron is a device for accelerating protons (or other charged particles) to high energies. The protons are held in a circular orbit of radius $R = p/eB$ by a uniform magnetic field B . Twice in each orbit they are subjected to an accelerating electric field. As they speed up, R increases and they spiral slowly outward. (a) Show that the period of each orbit is $T = 2\pi\gamma m/(eB)$. (b) As long as the motion is nonrelativistic, $\gamma \approx 1$ and the period is constant. This greatly simplifies the design of the cyclotron, since the accelerating field can be applied at a constant frequency. Assuming that a cyclotron can tolerate no more than a 2% increase in the period, what is the highest kinetic energy of the protons it can produce?

SECTION 2.10 (General Relativity *)

- 2.45** •• Using the argument sketched in Section 2.10 (the subsection on the gravitational redshift), show that the fractional frequency shift for light dropping through a height h is

$$\Delta f/f = gh/c^2 \quad (2.62)$$

- 2.46** •• One can get an estimate for the Schwarzschild radius using the following classical argument: Consider a spherical mass of radius R and mass M . (a) Using conservation of energy, write down the classical escape speed, the minimum speed v needed for an object to escape from the sphere's surface to infinity. The Schwarzschild radius R_S should be the value of R for which the escape speed equals the speed of light. Find an expression for R_S in terms of G , M , and c . (By a happy accident, this nonrelativistic argument produces the exact relativistic answer.) (b) What is R_S for a star of about 10 solar masses? (Solar mass $\approx 2 \times 10^{30}$ kg.). (c) Taking R_S as the radius of the black hole, what would be its average density?

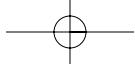
[Note: Your answer to part (a) should suggest that one could have black holes of any mass, however small. This is correct, but it turns out that such low-mass black holes would be extremely unstable and would evaporate very quickly.]

SECTION 2.11 (The Global Positioning System *)

- 2.47** •• Consider a GPS satellite in circular orbit 20,200 km above the earth's surface, moving with orbital speed v_s . If the speed of light were not constant, but instead obeyed the classical velocity-addition formula, then an observer on earth receiving a radio signal from a satellite would observe that the signal moves with speed $v_o = c \pm v_r$, where v_r is the radial speed of the satellite relative to the observer. (a) Find the orbital speed v_s of the satellite. (b) As a rough estimate, take $v_r \approx v_s$. (Computing v_r exactly is actually quite messy because of the complex relation of the velocities of the satellite and observer.) If the observer calculated the distance of the satellite using $r = v_o \Delta t$ instead of the correct $r = c \Delta t$, estimate the maximum error in r that would result.

- 2.48** •• There are many subtle effects that must be taken into account in the GPS calculations. Here you are to estimate just the effect of the time dilation of special relativity. (a) Find the speed of a GPS satellite, using the eight given in Fig. 2.12 and any other astronomical data (such as the earth's radius) you need from Appendix A. Hence find the time difference between a clock in a satellite and one on the ground after one complete orbit, assuming they were initially synchronized. (Ignore all effects except time dilation.) (b) Supposing we forgot to allow for this time difference, estimate the resulting error in the calculation of our position.

- 2.49** •• Do the same exercises as in Problem 2.48, but taking account of just the gravitational red shift instead of time dilation. [The difference in frequency of the two clocks is given by Eq. (2.62) in Problem 2.45, but



84 Chapter 2 • Relativistic Mechanics

in this problem g varies a lot over the height h , so in place of gh you should use $\int g \, dh$.]



COMPUTER PROBLEMS

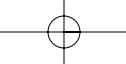
- 2.50** • (Section 2.4) If we can contrive to keep exerting a force on a particle (with a constant electric field, for example), we can increase its kinetic energy without limit. In nonrelativistic mechanics, this means that we can increase its speed without limit, but in relativity we know this is impossible. Find an expression for u in terms of K , and describe the behavior of u as K grows without limit. Use appropriate graphing software to plot u as a function of K for $0 \leq K \leq 4mc^2$.
- 2.51** •• (Section 2.7) A particle of mass m and charge q is released from rest in a uniform electric field \mathbf{E} directed along the x axis. Find its speed u as a function

of time. [Remember the definition (2.43) of force.] Now use graphing software to plot u as a function of t from $t = 0$ until $t = 10mc/q\mathbf{E}$.

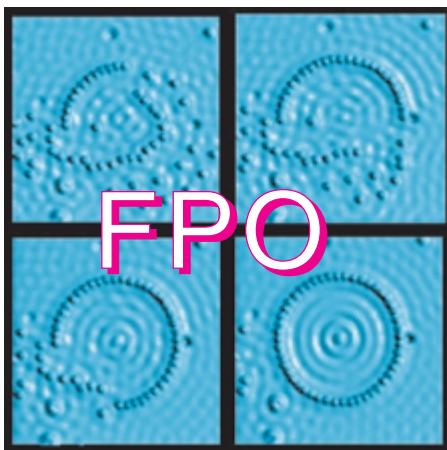
[*HINT:* When making this plot, you may as well assume that $m = c = q\mathbf{E} = 1$; this amounts to choosing your units in a convenient way.]

- 2.52** •• (Section 2.7) A particle of mass m and charge q is released from the origin in a uniform electric field \mathbf{E} directed along the x axis. Find its speed u as a function of x . [Remember the work-energy theorem (2.44).] Now use graphing software to plot u as a function of x from $x = 0$ to $x = 10mc^2/q\mathbf{E}$.

[*HINT:* When making this plot, you may as well assume that $m = c = q\mathbf{E} = 1$; this amounts to choosing your units in a convenient way.]



QUANTUM MECHANICS



- Chapter 3 Atoms
- Chapter 4 Quantization of Light
- Chapter 5 Quantization of Atomic Energy Levels
- Chapter 6 Matter Waves
- Chapter 7 The Schrödinger Equation in One Dimension
- Chapter 8 The Three-Dimensional Schrödinger Equation
- Chapter 9 Electron Spin
- Chapter 10 Multielectron Atoms: The Pauli Principle and the Periodic Table
- Chapter 11 Atomic Transitions and Radiation

In Part II (Chapters 3 to 11) we describe the second of the two theories that transformed twentieth-century physics — quantum theory. Just as relativity can be roughly characterized as the study of phenomena involving high speeds ($v \sim c$), so quantum theory can be described as the study of phenomena involving small objects — generally of atomic size or smaller.

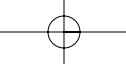
We begin our account of quantum theory in Chapter 3, with a descriptive survey of the main properties of atoms (their size, mass, constituents, and so on). In Chapters 4 and 5 we describe some puzzling properties of microscopic systems that began to emerge in the late nineteenth century — some properties of light in Chapter 4, and of atoms in Chapter 5. All of these puzzles pointed up the need for a new mechanics — quantum mechanics, as we now say — to replace classical mechanics in the treatment of microscopic systems. In Chapters 6 to 9 we describe the basic ideas of the new quantum mechanics, which began to develop around 1900 and was nearly complete by 1930. In particular, we introduce the Schrödinger equation, which is the basic equation of quantum mechanics, just as Newton's second law is the basic equation of Newtonian mechanics.

Armed with the Schrödinger equation, we can calculate most of the important properties of the simplest of all atoms, the hydrogen atom (with its one electron), as we describe in Chapter 8. In Chapter 9 we introduce one more important idea of the new quantum mechanics — the electron's spin angular momentum. Then in Chapters 10 and 11 we are ready to apply all these ideas to the general multielectron atom. We will see that quantum theory gives a remarkably complete account of all the 100 or so known different atoms, and that, in principle at least, it explains all of chemistry.

Although quantum mechanics was first developed to explain the properties of atoms, it has also been applied with extraordinary success to other systems, some larger than atoms, such as molecules and solids, and some smaller, such as subatomic particles. We will describe these other applications in Parts III and IV.

PART





Chapter 3

Atoms

- 3.1 Introduction
- 3.2 Elements, Atoms, and Molecules
- 3.3 Electrons, Protons, and Neutrons
- 3.4 Some Atomic Parameters
- 3.5 The Atomic Mass Unit
- 3.6 Avogadro's Number and the Mole
- 3.7 Kinetic Theory
- 3.8 The Mean Free Path and Diffusion
- 3.9 Brownian Motion
- 3.10 Thomson's Discovery of the Electron*
- 3.11 Millikan's Oil-Drop Experiment*
- 3.12 Rutherford and the Nuclear Atom*
- 3.13 Derivation of Rutherford's Formula*

Problems for Chapter 3

*These sections can be omitted without serious loss of continuity.

3.1 Introduction

Quantum physics is primarily the physics of microscopic systems. Historically, the most important such system was the atom, and even today the greatest triumph of quantum theory is the complete and accurate account that it gives us of atomic properties. For the next several chapters we will be discussing atoms, and in this chapter, therefore, we give a brief description of the atom and its constituents, the electron, proton, and neutron, together with some of the evidence for these entities.

3.2 Elements, Atoms, and Molecules

The concept of the atom arose in scholars' search to identify the basic constituents of matter. More than 2000 years ago Greek philosophers had recognized that this search requires answers to two questions: First, among the thousands of different substances we find around us — sand, air, water, soil, gold, diamonds — how many are the basic substances, or **elements**, from which all the rest are formed? Second, if one took a sample of one of these elements and subdivided it over and over again, could this process of repeated subdivision go on forever, or would one eventually arrive at some smallest, indivisible unit, or **atom** (from a Greek word meaning “indivisible”)? With almost no experimental data, the ancient Greeks could not find satisfactory answers to these questions; but this in no way lessens their achievement in identifying the right questions to ask.

Serious experimental efforts to identify the elements began in the eighteenth century with the work of Lavoisier, Priestley, and other chemists. By the end of the nineteenth century, about 80 of the elements had been correctly identified, including all of the examples listed in Table 3.1. Today we know that there are 90 elements that occur naturally on earth. All of the matter on earth is made up from these 90 elements, occasionally in the form of a pure element, but usually as a chemical compound of two or more elements, or as a mixture of such compounds.

TABLE 3.1

A few of the elements with their chemical symbols. For a complete list of the elements, see the periodic table inside the back cover or the alphabetical list in Appendix C.

Hydrogen (H)	Oxygen (O)	Iron (Fe)
Helium (He)	Sodium (Na)	Nitrogen (N)
Carbon (C)	Aluminum (Al)	Lead (Pb)
Copper (Cu)	Chlorine (Cl)	Uranium (U)

In addition to the 90 naturally occurring elements, there are about two dozen elements that can be created artificially in nuclear reactions. All of these artificial elements are unstable and disintegrate with half-lives much less than the age of the earth; this means that even if any of them were present when the earth was formed, they have long since decayed and are not found naturally in appreciable amounts.*

The evidence that the elements are composed of characteristic smallest units, or atoms, began to emerge about the year 1800. Chemists discovered the law of definite proportions: When two elements combine to form a pure chemical compound, they always combine in a definite proportion by mass. For example, when carbon (C) and oxygen (O) combine to form carbon monoxide (CO), they do so in the proportion 3:4; three grams of C combine with four of O to form seven grams of CO.



If we were to add some extra carbon, we would not get any additional CO; rather, we would get the same 7 g of CO, with all the extra carbon remaining unreacted.

The law of definite proportions was correctly interpreted by the English chemist John Dalton as evidence for the existence of atoms. Dalton argued that if we assume that carbon and oxygen are composed of atoms whose masses are in the ratio[†] 3:4, and if CO is the result of an exact pairing of these atoms (one atom of C paired with each atom of O), the law of definite proportions would immediately follow: The total masses of C and O that combine to form CO would be in the same ratio as the masses of the individual atoms, namely 3:4. If we add extra C (without any additional O), the extra C atoms have no O atoms with which to pair; thus we get no additional CO and the extra C remains unreacted.

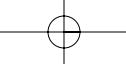
*A few elements with half-lives much less than the earth's age are observed to occur naturally, but this is because some natural nuclear process (radioactivity or cosmic-ray collisions, for example) is creating a fresh supply all the time.

[†]In stating Dalton's arguments we have replaced his terminology and measured masses with their modern counterparts. In particular, Dalton had measured the ratio of the C and O masses to be 5:7, not 3:4, as we now know it.

John Dalton (1766–1844, English chemist)



The son of an English weaver, Dalton left school at age eleven. A year later he became a schoolteacher, and this led to his interest in science. In his book, *New System of Chemical Philosophy*, Dalton argued that the chemists' law of definite proportions is evidence for the existence of atoms. Using this idea, he was the first to draw up a table of relative atomic masses. He had many other interests, including meteorology and color blindness. (He was himself color-blind — which can't have helped his chemical research.) He was a Quaker and extremely modest. In 1810 he declined membership in the Royal Society, although he was subsequently elected without his prior consent.



88 Chapter 3 • Atoms

We now know that this interpretation of the law of definite proportions was exactly correct. However, its general acceptance took 60 years or more, mainly because the situation was considerably more complicated than our single example suggests. For instance, carbon and oxygen can also combine in the ratio 3:8 to form carbon dioxide, CO_2 . Dalton was aware of this particular complication and argued (correctly) that the carbon atom must be able to combine with one O atom (to form CO) *or* two O atoms (to form CO_2). Nevertheless, this kind of ambiguity clouded the issue for many years.

A stable group of atoms, such as CO or CO_2 , is called a **molecule**. A few examples of molecules are listed in Table 3.2, where we follow the convention that a subscript on any symbol indicates the number of atoms of that kind (and single atoms carry no subscript — thus H_2O has two H atoms and one O atom). As the list shows, a molecule may contain two or more different atoms, for example, the CO and $\text{C}_6\text{H}_{12}\text{O}_6$ molecules, in which case we say that the two or more elements have combined to form a compound, of which the molecule is the smallest unit. Molecules can also contain atoms of just one kind, for example, O_2 . In this case, we do not speak of a new compound; we say simply that the element normally occurs (or sometimes occurs) not as separate atoms, but as groups of atoms clustered together, that is, as molecules.* Molecules can contain small numbers of atoms, like the examples in Table 3.2; but certain organic molecules, such as proteins, can contain tens of thousands of atoms.

TABLE 3.2

A few simple molecules.

Carbon monoxide, CO	Oxygen, O_2	Glucose, $\text{C}_6\text{H}_{12}\text{O}_6$
Water, H_2O	Nitrogen, N_2	Ethyl alcohol, $\text{C}_2\text{H}_6\text{O}$
Ammonia, NH_3	Sulfur, S_8	Urea, CON_2H_4

Although most of the credit for establishing the existence of atoms and molecules goes to the chemists, a second strand of evidence came from the kinetic theory of gases — which we would regard today as a part of physics, and which we discuss in sections 3.7 to 3.9. This theory assumes, correctly, that a gas consists of many tiny molecules in rapid motion. By applying Newton's laws of motion to the molecules, one could explain several important properties of gases (Boyle's law, viscosity, Brownian motion, and more). These successes gave strong support to the atomic and molecular hypotheses. In addition, kinetic theory, unlike the chemical line of reasoning, gave information on the actual size and mass of molecules. (The law of definite proportions implied that the C and O atoms have masses in the proportion 3:4 but gave no clue as to the actual magnitude of either.) Kinetic theory gave, for example, an expression for the viscosity of a gas in terms of the size of the individual molecules. Thus, measurement of viscosity allowed one to determine the actual size of the molecules — a method first exploited by Loschmidt in 1885.

By the beginning of the twentieth century it was fairly generally accepted that all matter was made up of elements, the smallest units of which were

*This possibility was probably the source of greatest confusion for the nineteenth-century chemists; if one wrongly identified the O_2 molecule as an atom, then one's interpretation of all other molecules containing oxygen — CO, CO_2 , and so on — would also be incorrect.

atoms. Atoms could group together into molecules, the formation of which explained the chemical compounds. The relative masses of many atoms and molecules were known quite accurately, and there were already reasonably reliable estimates of their actual masses.

3.3 Electrons, Protons, and Neutrons

Our story so far has the atom — true to its name — as the indivisible smallest unit of matter, and until the late nineteenth century there was, in fact, no direct evidence that atoms could be subdivided. The discovery of the first subatomic particle, the negatively charged electron, is generally attributed to J. J. Thomson (1897), whose experiments we describe briefly in Section 3.10. On the basis of his experiments, Thomson argued that electrons must be contained inside atoms, and hence that the atom is in fact divisible.

Fourteen years later (1911) Ernest Rutherford argued convincingly for the now familiar picture of the atom as a tiny planetary system, in which the negative electrons move far outside a positive **nucleus**. Rutherford's conclusion was based on an experiment that was the forerunner of many modern experiments in atomic and subatomic physics. In these experiments, called **scattering experiments**, one fires a subatomic projectile, such as an electron, at an atom or nucleus. By observing how the projectile is deflected, or *scattered*, one can deduce the properties of the target atom or nucleus. In the Rutherford experiment, the projectiles were alpha (α) particles — positively charged, subatomic particles ejected by certain radioactive substances. When these were directed at a thin metal foil, Rutherford found that almost all of them passed straight through, but that a few were deflected through large angles. In terms of his planetary model, Rutherford argued that the great majority of alpha particles never came close to any nuclei and encountered only a few electrons, which were too light to deflect them appreciably. On the other hand, a few of the alpha particles would pass close to a nucleus and would be deflected by the strong electrostatic force between the alpha particle and the nucleus. These, Rutherford argued, were the alpha particles that scattered through large angles. As we will describe in Section 3.12, Rutherford used his model to predict the number of alpha particles that should be scattered as a function of scattering angle and energy. The beautiful agreement between Rutherford's predictions and the experimental observations was strong evidence for the nuclear model of the atom, with its electrons far outside a tiny heavy nucleus.

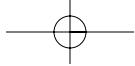
Within another eight years (1919) Rutherford had shown that the atomic nucleus can itself be subdivided, by establishing that nuclear collisions can break up a nitrogen nucleus. He identified one of the ejected particles as a hydrogen nucleus, for which he proposed the new name **proton** ("first one") in honor of its role in other nuclei. In 1932 Chadwick showed that nuclei contain a second kind of particle, the neutral **neutron**. (The experiments that identified the proton and neutron as constituents of nuclei will be described in Chapter 17.) With this discovery, the modern picture of the constituents of atoms was complete. Every atom contains a definite number of electrons, each with charge $-e$, in orbit around a central nucleus; and nuclei consist of two kinds of nuclear particles, or **nucleons**, the proton, with charge $+e$, and the uncharged neutron.

The constituents of five common atoms are shown in Table 3.3. In every case the number of electrons is equal to the number of protons, reflecting that the atom is neutral (in its normal state). Hydrogen is the only atom that has no

Joseph John Thomson (1856–1940, English)



As head of the famous Cavendish Laboratory in Cambridge for 35 years, J. J. Thomson was one of the most influential figures in the development of modern physics. Seven of his students and assistants went on to win the Nobel Prize, and he won it himself in 1906 for his discovery of the electron.



90 Chapter 3 • Atoms

Ernest Rutherford (1871–1937, British)



Rutherford grew up, one of twelve children, on a farm in New Zealand. In 1895 he won a scholarship to Cambridge, where he was a student of J. J. Thomson. After highly productive periods at McGill University in Montreal, Canada, and at Manchester University in England, he succeeded J. J. Thomson as Cavendish Professor at Cambridge in 1919. In 1908 he won the Nobel Prize in chemistry for showing that radioactivity transforms one element into another. Within another four years, he had established the existence of the atomic nucleus, and in 1917 he identified the first man-made nuclear reaction and proved that the proton is one of the constituents of nuclei. Rutherford's booming voice was known to disturb delicate experiments, and he is shown here (on the right) with his assistant Jack Ratcliffe under a sign that says "TALK SOFTLY PLEASE."

TABLE 3.3

The constituents of five representative atoms.

Atom	Electrons	Nucleons	
		Protons	Neutrons
Hydrogen, H	1	1	0
Helium, He	2	2	2
Carbon, C	6	6	6
Iron, Fe	26	26	30
Uranium, U	92	92	146

neutrons. In all other atoms the numbers of neutrons and protons are roughly equal. (We will see the reason for this in Chapter 16.) In many lighter atoms (helium or carbon, for example) the two numbers are exactly the same, while in most medium atoms the number of neutrons is a little larger; in the heaviest atoms there are about 50% more neutrons than protons.

In addition to their role as the building blocks of atoms, the electron and proton also define the smallest observed unit of charge. Their charges are exactly equal and opposite, with magnitude e ,

$$q_p = -q_e = e = 1.60 \times 10^{-19} \text{ C}$$

where "C" stands for coulomb. No charge smaller than e has been detected, and all known observed charges are integral multiples of e ,*

$$0, \pm e, \pm 2e, \pm 3e, \dots$$

Because e is so small, typical macroscopic charges are very large multiples of e , and the restriction to integral multiples is usually unimportant. On the atomic level, the existence of a smallest unit of charge is obviously very important.

By 1932 it appeared that all matter was made from just three subatomic particles, the electron, proton, and neutron. This picture of matter was a distinct simplification compared with its predecessor, with its 100 or so elements, each with a characteristic atom. As we will describe in Chapter 18, we now know that at least some of the subatomic particles themselves have an internal structure, being made of sub subatomic particles called quarks. However, for the purposes of atomic physics, and much of nuclear physics, the picture of matter as made of electrons, protons, and neutrons seems to be quite sufficient, and this is where we will stop the story for now.

3.4 Some Atomic Parameters

The distribution of electrons in their atomic orbits, and of protons and neutrons inside the nucleus, are two of the major concerns of quantum theory, as we will describe in several later chapters. Indeed, they are still the subjects of

*As we will discuss in Chapter 18, it has been established that certain "sub-subatomic" particles, called quarks, have charge $\pm e/3$ or $\pm 2e/3$. However, quarks are never found in isolation; they always occur as a member of a bound collection of quarks with total charge 0 or an integer multiple of e .

Section 3.4 • Some Atomic Parameters 91

current research. Nevertheless, a surprising number of atomic and nuclear properties can be understood just from an approximate knowledge of a few parameters, such as the size of the electron orbits and the masses of the electron, proton, and neutron. In this section we discuss some of these parameters, with which you should become familiar.

The size of an atom is not a precisely defined quantity, but can be characterized roughly as the radius of the outermost electron's orbit. This radius varies surprisingly little among the atoms, ranging from about 0.05 nm (helium) to about 0.3 nm (cesium, for example). In the majority of atoms it is between 0.1 and 0.2 nm. Thus for all atoms we can say that

$$\begin{aligned} \text{atomic radius} &\approx \text{radius of outer electron orbits} \\ &\sim 0.1 \text{ nm} = 10^{-10} \text{ m} \end{aligned} \quad (3.1)$$

The radii of nuclei are all much smaller than atomic radii and are usually measured in terms of the *femtometer* (fm) or fermi, defined as

$$1 \text{ fm} = 10^{-15} \text{ m} \quad (3.2)$$

Nuclear radii increase steadily from about 1 fm for the lightest nucleus, hydrogen (which is a single proton, of course), to about 8 fm for the heaviest nuclei (for example, uranium). Thus we can say that

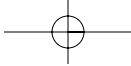
$$\text{nuclear radius} \approx \text{a few fm} = \text{a few} \times 10^{-15} \text{ m} \quad (3.3)$$

The single most important thing about the two sizes (3.1) and (3.3) is their great difference. The atomic radius is at least 10^4 times larger than the nuclear radius. Thus, if we made a scale model in which the nucleus was represented by a pea, the atom would be the size of a football stadium. Now, it turns out that in chemical reactions atoms approach one another only close enough for their outer electron orbits to overlap. This means that the chemical properties of atoms are determined almost entirely by the distribution of electrons, and are substantially independent of what is happening inside the nucleus. Further, quantum mechanics predicts (as we will find in Chapter 10) that the distribution of electrons is almost completely determined just by the *number* of electrons. Thus the chemical properties of an atom are determined almost entirely by the number of its electrons. This important number is called the **atomic number** and is denoted by the letter *Z*.

$$\begin{aligned} \text{atomic number, } Z, \text{ of an atom} \\ = \text{number of electrons in neutral atom} \\ = \text{number of protons in nucleus.} \end{aligned} \quad (3.4)$$

That *Z* is also the number of protons in the nucleus follows because the numbers of electrons and protons are equal in a neutral atom.

We should mention that many atoms can gain or lose a few electrons fairly easily. When this happens, we say that the atom is *ionized* and refer to the charged atom as an ion. (For example, the Ca^{2+} ion is a calcium atom that has lost two electrons; the Cl^- ion is a chlorine atom that has gained one electron.) Since the number of electrons in an atom can vary in this way, one must specify that *Z* is the number of electrons in the *neutral* atom, as was done in (3.4).



92 Chapter 3 • Atoms

Since chemical properties are determined mainly by the atomic number Z , one might expect that each chemical element could be identified by the atomic number of its atoms, and this proves to be so. All the atoms of a given element have the same atomic number Z , and, conversely, for every number Z between 1 and 109 (the highest value to be officially recognized with a name), there is exactly one element. Table 3.4 lists a few elements with their atomic numbers. We see that the hydrogen atom has one electron ($Z = 1$), helium two ($Z = 2$), carbon six, oxygen eight, and so on through uranium with 92, and on to the artificial Meitnerium with 109. An alphabetical list of all the elements can be found in Appendix C; a table of their properties — the periodic table — is inside the back cover.

TABLE 3.4

A few elements listed by atomic number, Z .

1 Hydrogen	6 Carbon	82 Lead
2 Helium	7 Nitrogen	⋮
3 Lithium	8 Oxygen	92 Uranium
4 Beryllium	⋮	⋮
5 Boron	26 Iron	109 Meitnerium

The mass of an atom depends on the masses of the electron, proton, and neutron, which are as follows:

$$\begin{aligned} \text{electron: } m_e &= 0.511 \text{ MeV}/c^2 = 9.11 \times 10^{-31} \text{ kg} \\ \text{proton: } m_p &= 938.3 \text{ MeV}/c^2 = 1.673 \times 10^{-27} \text{ kg} \\ \text{neutron: } m_n &= 939.6 \text{ MeV}/c^2 = 1.675 \times 10^{-27} \text{ kg} \end{aligned} \quad (3.5)$$

Within 1 part in 1000, the proton and neutron masses are equal, and, by comparison, the electron mass is negligible (one part in 2000). Thus we can say that

$$m_p \approx m_n \approx m_H \approx 940 \text{ MeV}/c^2 \quad (3.6)$$

where m_H denotes the mass of the hydrogen atom (a proton plus an electron). For many purposes it is sufficient to approximate the mass (3.6) as roughly $1 \text{ GeV}/c^2$.

The result (3.6) makes the approximate calculation of atomic masses extremely simple, since one has only to count the total number of nucleons (that is, the number of protons plus the number of neutrons). For example, the helium atom has two protons and two neutrons and so is four times as massive as hydrogen:

$$m_{\text{He}} \approx 4m_H$$

the carbon atom has six protons and six neutrons and so

$$m_C \approx 12m_H$$

Since its number of nucleons determines an atom's mass, this number is called the **mass number** of the atom. It is denoted by A :

$$\begin{aligned} \text{mass number, } A, \text{ of an atom} &= \text{number of nucleons in atom} \\ &= (\text{number of protons}) + (\text{number of neutrons}) \end{aligned} \quad (3.7)$$

An atom with mass number A has mass approximately equal to Am_H .

It often happens that two atoms with the same atomic number Z (and hence the same numbers of electrons and protons) have different numbers of neutrons in their nuclei. Such atoms are said to be **isotopes** of one another. Since the two isotopes have the same number of electrons, they have almost identical chemical properties and so belong to the same chemical element. But since they have different numbers of neutrons, their mass numbers, and hence masses, are different. For example, the commonest carbon atom has 6 protons and 6 neutrons in its nucleus, but there is also an isotope with 6 protons but 7 neutrons. These two atoms have the same chemical properties, but have different masses, about $12m_H$ and $13m_H$, respectively. To distinguish isotopes, we sometimes write the mass number A as a superscript before the chemical symbol. Thus the two isotopes of carbon just mentioned are denoted ^{12}C and ^{13}C (usually read as "carbon 12" and "carbon 13").

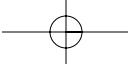
While some elements have only one type of stable atom, the majority have two or more stable isotopes. The maximum number belongs to tin with 10 stable isotopes, and on average there are about 2.5 stable isotopes for each element. Table 3.5 lists all of the stable isotopes of four representative elements.

TABLE 3.5

The stable isotopes of four elements. The percent of each element that occurs naturally in each isotope is shown in parentheses. A complete list, including all stable isotopes of all the elements and their abundances, is given in Appendix D.

Element	Atomic Number	Isotopes
Carbon	6	$^{12}\text{C}(98.9\%)$, $^{13}\text{C}(1.1\%)$
Magnesium	12	$^{14}\text{Mg}(79.0\%)$, $^{21}\text{Mg}(10.0\%)$, $^{26}\text{Mg}(11.0\%)$
Chlorine	17	$^{35}\text{Cl}(75.8\%)$, $^{37}\text{Cl}(24.2\%)$
Iron	26	$^{54}\text{Fe}(5.8\%)$, $^{56}\text{Fe}(91.7\%)$, $^{57}\text{Fe}(2.2\%)$, $^{58}\text{Fe}(0.3\%)$

Since the chemical properties of isotopes are so similar, the proportion of isotopes that occur in nature does not change in normal chemical processes. This means that the atomic mass of an element, as measured by chemists, is the weighted average of the masses of its various natural isotopes. This explains why chemical atomic masses are not always close to integer multiples of m_H . For example, we see from Table 3.5 that natural chlorine is $\frac{3}{4}$ the isotope ^{35}Cl and $\frac{1}{4}$ the isotope ^{37}Cl . Thus the chemical atomic mass of Cl is the weighted average $35.5m_H$. This complication caused some confusion in the historical development of atomic theory. The English physician Prout had pointed out as early as 1815 that the masses of atoms appeared to be integral multiples of m_H , suggesting that all atoms were made from hydrogen atoms. As more atomic masses were measured, examples of nonintegral masses (such as chlorine) were found, and Prout's hypothesis was rejected. Not until a hundred years later was it seen to be very nearly correct.



3.5 The Atomic Mass Unit

Since all atomic masses are approximately integer multiples of the hydrogen atom's mass, m_H , it would be natural to use a mass scale with m_H as the unit of mass, so that all atomic masses would be close to integers. In fact, this is approximately (although not exactly) how the atomic mass scale is defined, and the so-called **atomic mass unit**, or u , is to a good approximation just m_H .

To understand the exact definition of the atomic mass unit, we must examine why atomic masses are only approximately integral multiples of m_H . Let us consider an atom with atomic number Z and mass number A . The number of neutrons we denote by N , so that

$$A = Z + N$$

If we ignore for a moment the requirements of relativity, the mass of our atom would be just the sum of the masses of Z electrons, Z protons, and N neutrons.

$$m = Zm_e + Zm_p + Nm_n = Zm_H + Nm_n \quad (3.8)$$

To the extent that $m_n \approx m_H$ we can write this as

$$m \approx Zm_H + Nm_H \quad (3.9)$$

or

$$m \approx Am_H \quad (3.10)$$

In fact, of course, m_H and m_n are not exactly the same, the difference being of order 1 part in 1000. Thus the step from (3.8) to (3.9) is only an approximation, and our conclusion (3.10) can be in error by about 1 part in 1000.

There is a second, and more important, reason why (3.10) is only an approximation. We have learned from relativity that when bodies come together to form a stable bound system, the bound system has less mass than its separate constituents by an amount

$$\Delta m = \frac{B}{c^2} \quad (3.11)$$

where B is the binding energy, or energy required to pull the bound system apart into its separate pieces. For example, the helium nucleus (2 protons + 2 neutrons) has a binding energy of about 28 MeV; thus the helium atom has $28 \text{ MeV}/c^2$ less mass than the nearly $4000 \text{ MeV}/c^2$ predicted by (3.10). This is a correction of about 7 parts in 1000 and is appreciably more important than the correction discussed previously.*

For all atoms (except hydrogen itself) there is a similar adjustment due to the nuclear binding energy. Further, it turns out that in almost all atoms the correction is in the same proportion, namely 7 or 8 parts in 1000. (We will see

*There is also a correction for the binding energy of the electrons, but this is much smaller (1 part in 10,000 at the most) and is almost always completely negligible.

why the correction is always about the same in Chapter 16.) The only significant exception to this statement is the hydrogen atom itself, which has no nuclear binding energy and hence no correction. Thus if we want a mass scale on which most atomic masses are as close as possible to integers, the mass of the H atom is certainly not the best choice for the unit. If we chose instead ${}^4\text{He}$ or ${}^{12}\text{C}$ and defined the unit as $\frac{1}{4}$ the mass of ${}^4\text{He}$ or $\frac{1}{12}$ the mass of ${}^{12}\text{C}$, all atoms (except hydrogen) would have masses closer to integral multiples of the atomic mass unit. Which atom we actually choose is a matter of convenience, and by international agreement we use ${}^{12}\text{C}$, defining the u as follows:*

$$\begin{aligned} 1 \text{ atomic mass unit} &= 1 \text{ u} = \frac{1}{12} (\text{mass of one neutral } {}^{12}\text{C atom}) \\ &= 1.661 \times 10^{-27} \text{ kg} = 931.5 \text{ MeV}/c^2 \end{aligned} \quad (3.12)$$

With the definition (3.12) we can write the mass of an atom whose mass number is A as

$$(\text{mass of atom with mass number } A) \approx A \text{ u} \quad (3.13)$$

For most atoms this approximation is good to 1 part in 1000, and even in the worst case, hydrogen, it is better than 1 part in 100. Both of these claims are illustrated by the examples in Table 3.6.

TABLE 3.6

Atomic masses in u are always close to an integer.

${}^1\text{H}$	1.008	${}^{35}\text{Cl}$	34.969
${}^2\text{He}$	4.003	${}^{56}\text{Fe}$	55.939
${}^{12}\text{C}$	12 exactly	${}^{208}\text{Pb}$	207.977
${}^{16}\text{O}$	15.995	${}^{238}\text{U}$	238.049

Example 3.1

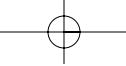
Using the natural abundances in Table 3.5, find the chemical atomic mass of magnesium in u to three significant figures.

The required mass is the weighted average of the masses of the isotopes ${}^{24}\text{Mg}$, ${}^{25}\text{Mg}$, ${}^{26}\text{Mg}$, as listed in Table 3.5. To three significant figures we can use the approximation (3.13) to give

$$\begin{aligned} \text{average atomic mass of Mg} \\ = [(0.79 \times 24) + (0.10 \times 25) + (0.11 \times 26)] \text{ u} = 24.3 \text{ u} \end{aligned} \quad (3.14)$$

in agreement with the observed value given inside the back cover. On those rare occasions when one needs greater accuracy, the more precise atomic masses given in Appendix D can be used.

*The atomic mass unit defined here, sometimes called the unified mass unit, replaces two older definitions, one based on ${}^{16}\text{O}$ and the other on the natural mixture of the three oxygen isotopes.



96 Chapter 3 • Atoms

3.6 Avogadro's Number and the Mole

From a fundamental point of view the natural way to measure a quantity of matter is to count the number of molecules (or atoms). In practice, a convenient amount of matter usually has a very large number of molecules, and it is sometimes better to have a larger unit than the individual molecule. The usual choice for that larger unit is the mole, which is defined as follows.

We first define **Avogadro's number**, N_A , as the number of atoms in 12 grams of ^{12}C .

$$\begin{aligned} \text{Avogadro's number, } N_A \\ = (\text{number of atoms in 12 grams of } ^{12}\text{C}) = 6.022 \times 10^{23} \end{aligned} \quad (3.15)$$

In this definition there is nothing sacred or fundamental about 12 grams; it is simply a reasonable macroscopic amount. (There is something *convenient* about the number 12 when used in conjunction with ^{12}C , as we will see directly.) Since N_A atoms of ^{12}C (each of mass 12 u) have total mass 12 g we see that

$$N_A \times (12 \text{ u}) = 12 \text{ grams}$$

Cancelling the 12 and dividing, we see that

$$N_A = \frac{1 \text{ gram}}{1 \text{ u}} \quad (3.16)$$

That is, N_A is just the number of u in a gram.

We now define a mole of objects (carbon atoms, water molecules, physicists) as N_A objects. Thus a mole of carbon is N_A carbon atoms, a mole of water is N_A water molecules (and a mole of physicists is N_A physicists). Since a mole always contains the same number of objects, the mass of a mole is proportional to the mass of the object concerned. From its definition, a mole of ^{12}C has mass exactly 12 grams. Thus a mole of ^4He has mass approximately 4 grams (more exactly 4.003 grams); a mole of hydrogen atoms, about 1 gram; a mole of H_2 molecules, about 2 grams; a mole of water, about 18 grams (since the H_2O molecule has mass $2 + 16 = 18$ u); and so on. The official abbreviation for the mole is "mol."

In chemistry the masses of atoms and molecules are usually specified by giving, not the mass of an individual atom or molecule, but the mass of a mole. Thus, instead of saying that

$$(\text{mass of } ^{12}\text{C atom}) = 12 \text{ u}$$

one says that

$$(\text{mass of 1 mole of } ^{12}\text{C}) = 12 \text{ grams}$$

Masses expressed in this way, in grams per mole, are often misleadingly called atomic or molecular weights. Evidently, the mass of an atom or molecule in u has the same numerical value as its atomic or molecular weight in grams per mole.

The mole is not a fundamental unit. It is an arbitrary unit, which is convenient for people who deal with large numbers of atoms and molecules. (It can be compared with the “dozen,” another arbitrary unit, which is convenient for people who deal with large numbers of eggs or doughnuts.) A purely theoretical physicist has little reason to work with either moles or Avogadro’s number, but for most of us, it is important to be able to translate between the fundamental language of molecules and the practical language of moles.

To conclude this brief discussion of the mole, we should mention that the mole is officially considered to have an independent dimension called the “amount of substance” (in the same way that meters have the dimension called “length”). For this reason Avogadro’s number,

$$N_A = 6.022 \times 10^{23} \text{ objects/mole}$$

is often called Avogadro’s *constant*, since it is viewed as a dimensional quantity with the units mole⁻¹ (or mol⁻¹).

3.7 Kinetic Theory

The description of the macroscopic properties of gases in terms of the microscopic behavior of atoms is called **kinetic theory**. In 1900 the notion that the world is made of very tiny particles, atoms, was widely, but not fully, accepted by the scientific community. As the twentieth century dawned, a vocal minority of competent scientists refused to accept the concept of atoms as fact because there was no direct, irrefutable evidence of their existence. Atoms were much too small to see directly; it was not until 1953 that individual atoms were first imaged with the newly invented field-ion microscope.* The evidence for the existence of atoms from chemistry, though viewed as compelling by many scientists, was indirect and not considered conclusive by all. Only when the increasingly precise predictions of kinetic theory were borne out by experiment did the scientific community completely embrace the atomic hypothesis. The last skeptics were finally convinced when experiment verified the theory of Brownian motion (described in Sec. 3.9) — a theory developed by a young, unknown, scientist named Albert Einstein.

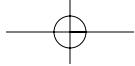
As our first application of kinetic theory, we show that the ideal gas law ($pV = nRT$), an experimentally discovered relation, can be derived from considerations of the motion of atoms. We will also find that the microscopic view of the world afforded by kinetic theory provides a very intuitive and satisfactory explanation of the meaning of temperature, a concept that was poorly understood before kinetic theory. Thus, the kinetic theory derivation of the ideal gas law was an important step in the path to the modern atomic view of matter.

Chemists usually write the ideal gas law as

$$pV = nRT \quad (3.17)$$

where n is the number of moles of the gas, the constant $R = 8.314 \text{ J/K} \cdot \text{mole}$ is called the **universal gas constant**, and p , V , and T are the pressure, volume, and

*This kind of microscope and others capable of resolving atoms are further described in Chapter 14.



98 Chapter 3 • Atoms

temperature of the gas. Physicists usually think in terms of the number N of molecules, rather than the number n of moles, so they replace n by $n = N/N_A$ and rewrite (3.17) as $pV = (N/N_A)RT$ or

$$pV = Nk_B T \quad (3.18)$$

where the constant k_B is called **Boltzmann's constant**.

$$k_B = \frac{R}{N_A} = 1.38 \times 10^{-23} \text{ J/K} = 8.62 \times 10^{-5} \text{ eV/K} \quad (3.19)$$

In 1900 the value of the universal gas constant was well established by comparing (3.17) with experiment. However, Avogadro's number N_A was not well known. Estimates of its value varied by a factor of 100 or more. Basically, there was no good way known to count atoms. The inability of experimentalists to determine Avogadro's number contributed to skeptics' arguments against the atomic hypothesis.

We now derive the ideal gas law in the form $pV = NkT$ by considering the motion of atoms. We consider a gas of N atoms or molecules, each with mass m , in **thermal equilibrium** enclosed in a rectangular container of length L and cross-sectional area A . A system is in thermal equilibrium when all parts of the system have the same average energy. We assume that the gas is so dilute and the interactions among the molecules are so weak that the gas can be considered ideal, that is, the molecules are noninteracting. Our goal is to derive an expression for the pressure that the gas exerts on the walls of the container, in terms of the microscopic properties of the molecules.

We orient a coordinate system so that the right and left walls of the container are perpendicular to the x axis as shown in Fig. 3.1. Consider now a particular molecule labeled i moving toward the right wall with an x component of velocity $+v_{xi}$. If the wall is smooth and rigid, the molecule will bounce off the wall like a light ray reflecting from a mirror, and after the collision, the new velocity will have x component $-v_{xi}$ while the y and z components will be unchanged. The change in momentum of the molecule is then $\Delta p_{xi} = p_x^{final} - p_x^{initial} = -2mv_{xi}$. The time Δt_i between collisions of molecule i with the right side of the container is the time for the molecule to bounce back and forth once between the right and left walls, so $\Delta t_i = 2L/v_{xi}$. Recall that we are assuming a gas of non-interacting atoms, so that each molecule bounces between the right and left walls without interacting with the other molecules. Except for extremely dilute gases, this assumption of *complete* non-interaction is unfounded, but, as we will discuss below, collisions among the atoms do not affect our final results. The force from this one molecule on the right wall is a series of brief impulses, but the *time-averaged* force due to this molecule is the total change in momentum between collisions divided by the total time, (recall that $\mathbf{F} = \Delta \mathbf{p}/\Delta t$)

$$F_{xi} = \frac{|\Delta p_{xi}|}{\Delta t_i} = \frac{|2mv_{xi}|}{\Delta t_i} = \frac{mv_{xi}^2}{L}$$

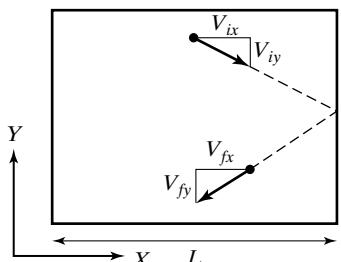


FIGURE 3.1

A gas molecule bounces off the wall of a container.

The total pressure p on the wall due to all the molecules in the container is obtained by summing over all N molecules and dividing by the area A of the wall.

$$p = \frac{\sum_i F_{xi}}{A} = \frac{m \sum_i v_{xi}^2}{AL} = \frac{Nm}{V} \frac{\sum_i v_{xi}^2}{N} \quad (3.20)$$

In the last step we have replaced AL with the volume V of the box, and we have multiplied and divided by N . The last expression $\sum v_{xi}^2/N = \langle v_x^2 \rangle$ is the average value of v_x^2 . (The brackets $\langle \dots \rangle$ represent an average over all molecules). Equation (3.20) now becomes

$$p = \frac{Nm\langle v_x^2 \rangle}{V} \quad (3.21)$$

Finally, we relate $\langle v_x^2 \rangle$ to the average squared speed of a molecule $\langle v^2 \rangle$. Because there is no preferred direction for the motion of the molecules, we must have

$$\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle \quad (3.22)$$

The square of the speed of a molecule is $v^2 = v_x^2 + v_y^2 + v_z^2$. Therefore,

$$\langle v^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle = 3\langle v_x^2 \rangle \quad (3.23)$$

and equation (3.21) becomes

$$pV = Nm\frac{1}{3}\langle v^2 \rangle \quad (3.24)$$

a result obtained by the Swiss mathematician Daniel Bernoulli at the amazingly early date of 1760.

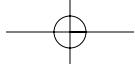
Comparing this expression with the empirical relation

$$pV = NkT$$

we see that our derived expression matches the ideal gas law if we assume that the average kinetic energy of a molecule $\langle KE \rangle = \frac{1}{2}m\langle v^2 \rangle$ is related to the absolute temperature T by

$$\frac{1}{2}m\langle v^2 \rangle = \frac{3}{2}kT \quad (3.25)$$

Our derivation has not only predicted the form of the ideal gas law but has also shown the physical meaning of temperature. Equation (3.25) says that temperature is proportional to the average kinetic energy per molecule. For a gas, “hotter” means “more kinetic energy per molecule, faster moving molecules.” In Chapter 15 we will give a more general definition of temperature, one that applies to solids as well as gases, but the basic idea remains the same: *Temperature is a measure of the energy per atom.*



100 Chapter 3 • Atoms

3.8 The Mean Free Path and Diffusion

In this section we examine some further features of the kinetic theory of gases. In particular, we look at the consequences of the fact that gases are made of molecules that have a nonzero size. In the derivation of the ideal gas law, we made the assumption that molecules collide only with the walls of the container and not with each other. In fact, because of their nonzero size, molecules do collide with each other, but these intermolecule collisions* alter neither the average speed of the molecules nor the validity of equation (3.21), so the ideal gas law remains valid in the presence of collisions. [An alternate derivation of (3.21), in the presence of collisions, is explored in Problem 3.36.]

The average distance that a molecule travels between collisions with other molecules is called the **mean free path**. As one might expect, the mean free path depends on both the size and density of the gas molecules; collisions are more frequent if the molecules are larger and if there are more of them. If we approximate the molecules as spheres of radius R , two molecules will overlap and hence collide if their centers come within $2R$ of each other (see Fig. 3.2). The **collision cross section** is defined as the collision area $\sigma = \pi(2R)^2$ that a molecule presents as it moves through space. We now show that the mean free path λ , the collision cross section σ , and the number density n of molecules (n = number per volume[†]) are related by the simple equation

$$\lambda \approx \frac{1}{n\sigma} \quad (3.26)$$

Consider a molecule that undergoes a large number N of collisions and in so doing moves along a zigzag path of total length L (see Fig. 3.3). Since λ is defined as the mean distance traveled between collisions, we have $L = N\lambda$. As the molecule travels the distance L , it sweeps through a volume of length L and cross-sectional area σ , colliding with the N molecules it encounters whose centers are within this volume $V = L\sigma = N\lambda\sigma$. Assuming that the number density N/V in this zigzag tube-shaped volume is the same as the number density n in the sample as a whole, we have

$$n = \frac{N}{V} = \frac{N}{N\lambda\sigma} = \frac{1}{\lambda\sigma}$$

which gives the desired result (3.26). This derivation is not rigorously correct because we have not taken proper account of the fact that the other molecules are moving. A more careful calculation yields $\lambda = 1/(\sqrt{2}n\sigma)$.

Example 3.2

Given that the mean radius of an air molecule (either O_2 or N_2) is about $R = 0.15$ nm, what is the approximate mean free path λ of the molecules in air at atmospheric pressure and room temperature? (1 atm $\approx 1.01 \times 10^5$ Pa)

*We are assuming so-called hard-sphere collisions, that is, a short-range repulsive interaction only, no long-range interaction.

[†]Note well that, here, n is *not* the number of moles, as in (3.17). It is the number of molecules per unit volume.

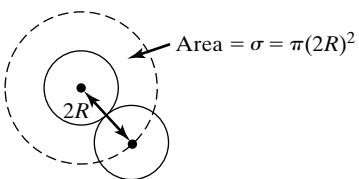


FIGURE 3.2

Two molecules will collide if their centers are closer than $2R$, where R is the radius of either molecule.

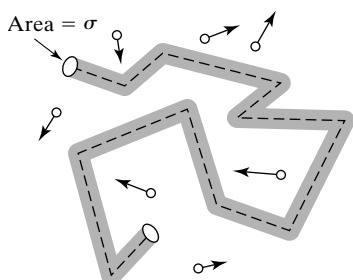


FIGURE 3.3

A molecule in gas collides with its neighbors, sweeping out a tube with cross-sectional area σ . The mean free path λ is the average length of the separate straight sections.

Section 3.8 • The Mean Free Path and Diffusion 101

To use the formula $\lambda \approx 1/(n\sigma)$, we must compute the number density n and the cross section σ . From the ideal gas law (3.18), we have

$$n = \frac{N}{V} = \frac{P}{kT} = \frac{1.01 \times 10^5 \text{ Pa}}{(1.38 \times 10^{-23} \text{ J/K}) \times (300 \text{ K})} = 2.45 \times 10^{25} \text{ m}^{-3}$$

The collision cross section is

$$\sigma = \pi(2R)^2 = 4\pi(0.15 \times 10^{-9} \text{ m})^2 = 2.8 \times 10^{-19} \text{ m}^2$$

So we have

$$\lambda \approx \frac{1}{n\sigma} = \frac{1}{(2.45 \times 10^{25}) \times (2.8 \times 10^{-19})} = 1.4 \times 10^{-7} \text{ m} = 140 \text{ nm}$$

This is about 40 times greater than the average distance between nearest-neighbor molecules in the gas.

The convoluted zigzag motion of an individual molecule as it collides with its neighbors is an example of **diffusive motion** or **diffusion**. We will show next that with diffusion the average magnitude of the displacement D attained by a molecule in a time t is proportional to $t^{1/2}$,

$$D_{\text{dif}} \propto \sqrt{t}$$

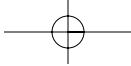
This is to be contrasted with ordinary **ballistic motion**, which is the straight-line, constant-velocity motion of a projectile in the absence of friction or gravity. For ballistic motion, with speed v , the net displacement is $D = vt$; that is, displacement is proportional to time,

$$D_{\text{bal}} \propto t$$

Compared to a ballistic projectile, a diffuser with the same speed does not get very far.

Diffusion is also called **random-walk motion** because the motion is that of a bewildered walker who wanders about, making frequent, random changes in direction. The standard random-walk problem is usually formulated like this: A walker moving in 1, 2, or 3 dimensions, takes a series of steps labeled $i = 1, 2, \dots, N$. The i^{th} step is described by a vector \mathbf{d}_i , which has length d , and a *random* direction. After N steps, the net displacement of the walker, relative to his starting point, is

$$\mathbf{D} = \sum_{i=1}^N \mathbf{d}_i \quad (3.27)$$



102 Chapter 3 • Atoms

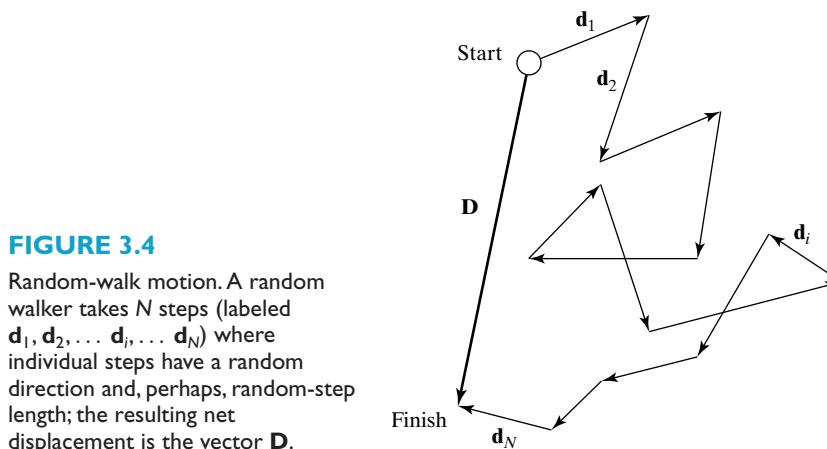


FIGURE 3.4

Random-walk motion. A random walker takes N steps (labeled $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_i, \dots, \mathbf{d}_N$) where individual steps have a random direction and, perhaps, random-step length; the resulting net displacement is the vector \mathbf{D} .

In the case of a gas, the walker is a molecule moving in 3 dimensions, taking steps of average length λ , the mean free path (see Fig. 3.4).

We seek the average distance of the walker from the origin after N steps. The word “average” must be carefully defined since there is more than one kind of average. We will use the **root-mean-square** average distance, which is defined as the square root of the average of the distance squared,

$$D_{\text{rms}} = \sqrt{\langle D^2 \rangle} \quad (3.28)$$

Here the brackets $\langle \dots \rangle$ indicate an average over all possible N -step random walks, often called the **ensemble average**. The rms average is usually chosen over other possible averages (such as the average of the absolute value of the displacement, $\langle |\mathbf{D}| \rangle$) because it is much easier to work with mathematically.

We now show that after N steps, each of length d , the rms average distance of a random walker from the starting point is

$$D_{\text{rms}} = \sqrt{Nd} \quad (3.29)$$

We begin by noting that the magnitude of the vector \mathbf{D} is $D = |\mathbf{D}| = \sqrt{\mathbf{D} \cdot \mathbf{D}}$. The dot product of \mathbf{D} with itself is

$$\mathbf{D} \cdot \mathbf{D} = D^2 = \left(\sum_{i=1}^N \mathbf{d}_i \right) \cdot \left(\sum_{j=1}^N \mathbf{d}_j \right) = \underbrace{\sum_{i=1}^N \mathbf{d}_i^2}_{Nd^2} + \sum_{i \neq j} \mathbf{d}_i \cdot \mathbf{d}_j \quad (3.30)$$

We now take the ensemble average, the average over all possible N -step walks, and obtain

$$\langle D^2 \rangle = \langle Nd^2 \rangle + \underbrace{\left\langle \sum_{i \neq j} \mathbf{d}_i \cdot \mathbf{d}_j \right\rangle}_0 = Nd^2 \quad (3.31)$$

Since the direction of \mathbf{d}_i is random, the quantity $\mathbf{d}_i \cdot \mathbf{d}_j (i \neq j)$ is positive as often as it is negative, and so $\langle \sum \mathbf{d}_i \cdot \mathbf{d}_j \rangle$ averages to zero. Taking the square root of (3.31), we have our desired result

$$\sqrt{\langle D^2 \rangle} = D_{\text{rms}} = \sqrt{Nd}$$

Example 3.3

An aimless physics student performs a random walk by taking 1-meter long steps, at a rate of one step each second, with each step in a completely random direction. How long does it take, on average, for the student to wander to a point 1 kilometer from his starting point?

Using equation (3.29), with step length $d = 1 \text{ m}$ and final displacement $D_{\text{rms}} = 1000 \text{ m}$, we have for the number of steps N ,

$$N = \left(\frac{D_{\text{rms}}}{d} \right)^2 = \left(\frac{1000}{1} \right)^2 = 10^6$$

That is, on average, the student needs to take one million steps in order to get one kilometer away from his starting point. At 1 second per step, this walk will take nearly 12 days.

We can now derive a formula for how far a molecule has diffused in a gas after a time t . This is a random-walk problem in which the step length is the mean free path λ of the molecule. The rms average speed of our molecular random walker is, according to (3.25),

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}} \quad (3.32)$$

We can define a **collision time**, τ , as the mean time between collisions (that is, between steps) by the equation

$$v_{\text{rms}} = \frac{\lambda}{\tau} \quad (3.33)$$

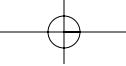
The time t required to make N steps is then $t = N\tau$. We can now rewrite (3.29) in terms of the time t of the walk and other microscopic variables

$$D_{\text{rms}} = \sqrt{N}\lambda = \sqrt{\frac{t}{\tau}}\lambda = \sqrt{\frac{v_{\text{rms}}t}{\lambda}}\lambda = \sqrt{\lambda v_{\text{rms}}} \sqrt{t}$$

Finally, using (3.26) and (3.32), we have

$$D_{\text{rms}} \approx \frac{1}{\sqrt{n\sigma}} \left(\frac{3kT}{m} \right)^{1/4} \sqrt{t} \quad (3.34)$$

This is a very useful formula for a present-day experimental physicist. However, to a physicist working in the late 1800s, it was interesting, but not especially useful, since none of the quantities σ , m , or D_{rms} could be measured experimentally. (Recall that at that time there was not even any good way to measure Avogadro's number.) However, in 1905 Einstein derived a similar formula for the case of larger particles, particles suspended in air or water and large enough to be seen and measured in a microscope.



104 Chapter 3 • Atoms

Example 3.4

A bottle of ammonia (NH_3) is opened in a closed room where the air is perfectly still. Estimate how long it will take for the evaporating ammonia molecules to diffuse from one side of the room to the other, a distance of 5 m. The mean radius of an ammonia molecule is about $R = 0.15 \text{ nm}$.

The size of the ammonia molecule is very nearly the same as that of an O_2 or N_2 molecule, so we can use the values of σ , the collision cross section, and n , the number density of molecules, from Example 3.2 above. The mass of an ammonia molecule is

$$\begin{aligned}m &= (\text{mass of 1 mole})/(\text{Avogadro's number}) = (17 \text{ g})/(6 \times 10^{23}) \\&= 2.8 \times 10^{-26} \text{ kg}\end{aligned}$$

Solving equation (3.34) for the time t , we have

$$\begin{aligned}t &\approx (D_{\text{rms}})^2 n \sigma \left(\frac{m}{3kT} \right)^{1/2} \\&= 5^2 \times (2.5 \times 10^{25}) \times (2.8 \times 10^{-19}) \times \left(\frac{2.8 \times 10^{-26}}{3 \times (1.38 \times 10^{-23}) \times (300)} \right)^{1/2} \\&\approx 3 \times 10^5 \text{ s} \approx 3 \text{ days}\end{aligned}\tag{3.35}$$

This long time seems to contradict everyday experience. If you open a bottle of ammonia, the smell will usually permeate the room in minutes or perhaps, at most, an hour. That is because, under normal circumstances, the mixing of air (indoors or out) is accomplished primarily by convection, not diffusion. The point here is that diffusion is very slow compared to convection; if you want to be sure that gases or liquids are mixed, you must stir them.

3.9 Brownian Motion

In 1828 a Scottish botanist, Robert Brown, discovered that tiny pollen grains, when suspended in water and viewed under a microscope, exhibited an irregular jiggling motion, which was later dubbed **Brownian motion** (see Fig. 3.5). Brown initially thought that this motion was due to some “life force”; however, he quickly established that tiny particles of any composition, whether organic or inorganic, suspended in any fluid, whether liquid or gas, also exhibited this erratic motion. In the decades that followed, Brownian motion was carefully studied and a variety of explanations were suggested. In the late 1880s, it was proposed that Brownian motion is caused by the random motion of the molecules in the surrounding fluid. A small suspended particle experiences a continuous, irregular pounding from all sides because of collisions with the molecules of the fluid. This explanation, which presumed the existence of atoms, was controversial; many thought Brownian motion was due to convection currents or vibrations transmitted through the fluid — explanations that did not require the existence of atoms.

In 1905 the twenty-six-year-old Einstein, having failed to procure an academic position, was supporting himself and his young family by working in a Swiss Patent Office as a “technical expert, third-class.” Though not very prestigious, this job brought him financial security and allowed him enough leisure

**FIGURE 3.5**

The positions of a Brownian particle recorded at regularly spaced time intervals, from the book *Atoms* by Jean Perrin.

time to pursue his own studies of physics. In that year, working in his spare time, Einstein wrote six history-making papers — a creative outburst rivaled only by the work of the young Isaac Newton, two and a half centuries before. The six articles, all published in the German scientific journal *Annalen der Physik* (Annals of Physics) consisted of a paper on the photoelectric effect (Section 4.3), for which he would receive the Nobel Prize in 1922; two papers on special relativity, the second of which contained the formula $E = mc^2$; a paper on the viscosity of dilute solutions, which turned out to have enormous practical importance and which is the Einstein paper most cited in the scientific literature; and two papers on the theory of Brownian motion.

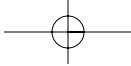
Einstein was able to derive an equation for the motion of a Brownian particle, an equation similar to (3.34), but one containing quantities that could be measured experimentally. Einstein showed that the average displacement D_{rms} in a time t of a spherical particle* of radius a , suspended in a fluid with **viscosity** η , is given by

$$D_{\text{rms}} = \sqrt{\frac{kT}{3\pi\eta a}} \sqrt{t} \quad (3.36)$$

(The viscosity of a fluid is a measure of its resistance to flow: thicker, more syrupy fluids have higher viscosity.)

In a series of painstaking measurements, the French experimentalist Jean-Baptiste Perrin verified Eq. (3.36) and used it to determine experimentally Boltzmann's constant k_B , and hence, Avogadro's number $N_A = R/k_B$. Perrin also determined N_A experimentally using two other methods: One method relied on a theory of the rotation of Brownian particles, a theory also due to Einstein; the other involved determining the distribution of suspended

*The Brownian particle is assumed to be much larger than an atom, but small enough to remain suspended in the fluid. The Brownian motion of larger-than-atom-sized particles is similar to, but not the same as the random-walk motion of individual atoms; the Brownian motion equation (3.36) cannot be derived from the atom motion equation (3.34).



106 Chapter 3 • Atoms

Jean Baptiste Perrin
(1870–1942, French)



Perrin was professor of chemistry at the University of Paris, but is best remembered for two important contributions to modern physics. His investigations of cathode rays, which he showed conclusively to carry negative charge, preceded by a couple of years Thomson's proof that these "rays" were actually the negative particles that we now call electrons. In 1913 Perrin published the results of his extensive studies of Brownian motion, which yielded a precise value for Avogadro's number, provided convincing direct proof of the atomic hypothesis, and earned him the Nobel Prize for physics in 1926.

particles in a gravitational field. All three methods produced the same value of N_A within a few percent. It was this concordance of different measurements of Avogadro's number that convinced almost all* the remaining skeptics of the atomic hypothesis.

3.10 Thomson's Discovery of the Electron*

* In the last four sections of this chapter, we describe three of the pivotal experiments of modern physics: the discovery of the electron, the measurement of its charge, and the discovery of the atomic nucleus. We believe that every well-educated scientist should know some of the details of these experiments. Nevertheless, we will not be using this material again, so if you are pressed for time, you could omit these sections without loss of continuity.

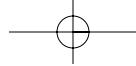
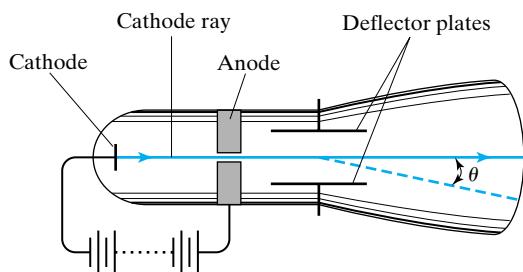
The main purpose of this chapter has been to introduce the principal characters in the story of atomic physics — atoms and molecules, electrons, protons, and neutrons — and to describe one of the first successes of atomic theory, the kinetic theory of gases. This early history, culminating in the acceptance of the atomic hypothesis by the scientific community, is not regarded as part of "modern physics," and in this book we will content ourselves with the brief description already given.[†] On the other hand, the more recent history of atomic physics is very much a part of modern physics, and from time to time we will give more detailed descriptions of some of its experimental highlights. In particular, we conclude this chapter by describing three key experiments: the experiments of J. J. Thomson and Millikan, which together identified the electron and its principal properties, and the so-called Rutherford scattering experiment, which established the existence of the atomic nucleus.

The discovery of the electron is generally attributed to J. J. Thomson (1897) for a series of experiments in which he showed that "cathode rays" were in fact a stream of the negative particles that we now call electrons. Cathode rays, which had been discovered some 30 years earlier, were the "rays" emitted from the cathode, or negative electrode, of a cathode ray tube — a sealed glass tube containing two electrodes and low-pressure gas (Fig. 3.6). When a large potential is applied between the cathode and anode, some of the gas atoms ionize and an electric discharge occurs. Positive ions hitting the cathode eject electrons, which are then accelerated toward the anode. With the arrangement shown in Fig. 3.6, some of the electrons pass through the hole in the anode and coast on to the far end of the tube. At certain pressures the rays can be seen by the glow that they produce in the gas, while at lower pressures they produce a fluorescent patch where they strike the end of the tube (as in a standard television tube).

It had been shown by William Crookes (1879) and others that cathode rays normally travel in straight lines and that they carry momentum. (When directed at the mica vanes of a tiny "windmill," they caused the vanes to rotate.) Crookes had also shown that the rays can be bent by a magnetic field, the direction of deflection being what would be expected for negative charges. All of this suggested that the rays were actually material particles carrying

*Not quite all. For instance, Ernst Mach (of Mach's number) died in 1916, still refusing to accept the reality of atoms.

[†]You can find an authoritative but highly readable history of atomic theory from about 600 B.C. to 1960 in *The World of the Atom* by H. A. Boorse and L. Motz, New York: Basic Books, 1966.

Section 3.10 • Thomson's Discovery of the Electron **107****FIGURE 3.6**

Thomson's cathode ray tube. When opposite charges were placed on the deflector plates, the "rays" were deflected through an angle θ as shown.

negative charge. Unluckily, attempts to deflect the rays by a transverse electric field had failed, and this failure had led to the suggestion that cathode rays were not particles at all, but were instead some totally new phenomenon — possibly some kind of disturbance of the ether.

J. J. Thomson carried out a series of experiments that settled these questions beyond any reasonable doubt. He showed that if the rays were deflected into an insulated metal cup, the cup became negatively charged; and that as soon as the rays were deflected away from the cup, the charging stopped. He showed, by putting opposite charges on the two deflecting plates in Fig. 3.6, that cathode rays were deflected by a transverse electric field. (He was also able to explain the earlier failures to observe this effect. Previous experimenters had been unable to achieve as good a vacuum as Thomson's. The remaining gas in the tube was ionized by the cathode rays, and the ions were attracted to the deflecting plates, neutralizing the charge on the plates and canceling the electric field.) After further experiments with magnetic deflection of cathode rays, Thomson concluded, "I can see no escape from the conclusion that they are charges of negative electricity carried by particles of matter." Accepting this hypothesis, he next measured some of the particles' properties. By making each measurement in several ways, he was able to demonstrate the consistency of his measurements and to add weight to his identification of cathode rays as negatively charged particles.

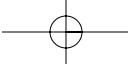
Thomson measured the speed of the cathode rays by applying electric and magnetic fields. The force on each electron in fields \mathbf{E} and \mathbf{B} is the well known Lorentz force

$$\mathbf{F} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (3.37)$$

(where \mathbf{v} is the electron's velocity). Using an E field alone, he measured the deflection of the electrons. After removing E and switching on a transverse magnetic field B , he adjusted B until the magnetic deflection was equal to the previous electric deflection. Under these conditions $\mathbf{E} = \mathbf{v} \times \mathbf{B}$, or since \mathbf{v} and \mathbf{B} were perpendicular,

$$v = \frac{E}{B} \quad (3.38)$$

(In Thomson's experiments v was of order $0.1c$. Given his experimental uncertainties, this means that his experiments can be analyzed nonrelativistically.) By measuring the heating of a solid body onto which he directed the electrons, he got a second estimate of v , which agreed with the value given by (3.38) considering his fairly large uncertainties. (See Problem 3.43.)



108 Chapter 3 • Atoms

Knowing the electrons' speed he could next find their "mass-to-charge" ratio, m/e . For example, as we saw in Chapter 2 — Eq. (2.48) — a B field causes electrons to move in a circular path of radius

$$R = \frac{mv}{eB} \quad (3.39)$$

(In discussing relativity, we used v for the relative speed of two frames and u for the speed of a particle. Here we have reverted to the more standard v for the electron's speed.) Thus, measurement of R (plus knowledge of v and B), gave the ratio m/e . Similarly, you can show (Problem 3.40) that an E field alone deflects the electrons through an angle

$$\theta = \frac{eEl}{mv^2} \quad (3.40)$$

where l denotes the length of the plates that produce E . Thus, by measuring the deflection θ produced by a known E field, he could again determine the ratio m/e .

Even though Thomson used several different gases in his tube and different metals for his cathode, he always found the same value for m/e (within his experimental uncertainties). From this observation, he argued correctly that there was apparently just one kind of electron, which must be contained in all atoms. He could also compare his value of m/e for electrons with the known values of m/e for ionized atoms. (The mass to charge ratio for ionized atoms had been known for some time from experiments in electrolysis — the conduction of currents through liquids by transport of ionized atoms.) Thomson found that even for the lightest atom (hydrogen) the value of m/e was about 2000 times greater than its value for the electron. The smallness of m/e for electrons had to be due to the smallness of their mass or the largeness of their charge, or some combination of both. Thomson argued (again correctly, as we now know) that it was probably due to the smallness of the electrons' mass.

3.11 Millikan's Oil Drop Experiment*

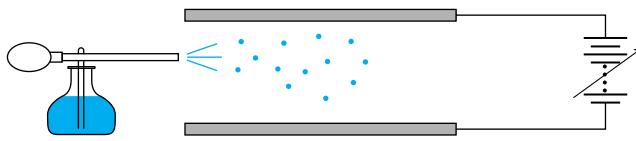
A frustrating shortcoming of Thomson's measurements was that they allowed him to calculate the ratio m/e , but not m or e separately. The reason for this is easy to see: Newton's second law for an electron in E and B fields states that

$$m\mathbf{a} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

Evidently, any calculation of the electron's motion is bound to involve just the ratio m/e and not m or e separately. One way around this difficulty was to study the motion of some larger body, such as a droplet of water, whose mass M could be measured and which had been charged by the gain or loss of an electron. In this way one could measure the ratio M/e . Knowing the mass M of the droplet, one could find e , and thence the mass m of the electron from the known value of m/e .

The need to measure e (or m) separately was recognized by Thomson and his associates, who quickly set about measuring e using drops of water. However, the method proved difficult (for example, the water drops evaporated rapidly), and Thomson could not reduce his uncertainties below about 50%.

Section 3.11 • Millikan's Oil Drop Experiment 109

**FIGURE 3.7**

Millikan's oil-drop experiment. The potential difference between the plates is adjustable, both in magnitude and direction.

In an effort to avoid the problem of evaporation, the American physicist Robert Millikan tried using oil drops and quickly developed an extremely effective technique. His apparatus is sketched in Fig. 3.7. Droplets of oil from a fine spray were allowed to drift into the space between two horizontal plates. The plates were connected to an adjustable voltage, which produced a vertical electric field E between the plates. With the field off, the droplets all drifted downward and quickly acquired their terminal speed, with their weight balanced by the viscous drag of the air. When the E field was switched on, Millikan found that some of the drops moved down more rapidly, while others started moving upward. This showed that the drops had already acquired electric charges of both signs, presumably as a result of friction in the sprayer. The method by which Millikan measured these charges was ingenious and intricate, but the essential point can be understood from the following simplified account. (For some more details, see Problem 3.45.)

By adjusting the electric field E between the plates, Millikan could hold any chosen droplet stationary. When this was the case, the upward electric force must exactly have balanced the downward force of gravity:

$$qE = Mg \quad (3.41)$$

where q was the charge on the droplet. Since E and g were known, it remained only to find M in order to determine the charge q .

To measure M , Millikan switched off the E field and observed the terminal speed v with which the droplet fell. (This speed was very small and so easily measured.) From classical fluid mechanics, it was known that the terminal speed of a small sphere acted on by a constant force F is

$$v = \frac{F}{6\pi r\eta} \quad (3.42)$$

where r is the radius of the sphere and η the viscosity of the gas in which it moves. In the present case F is just the weight of the droplet,

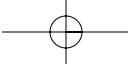
$$F = Mg = \frac{4}{3}\pi r^3 \rho g \quad (3.43)$$

where ρ is the density of the oil. Substitution of (3.43) in (3.42) gives the speed of the falling droplet as

$$v = \frac{2r^2 \rho g}{9\eta} \quad (3.44)$$

Since ρ , g , and η were known, measurement of v allowed Millikan to calculate the droplet's radius r and thence, from (3.43), its mass M . Using (3.41), he could then find the charge q .

From time to time it was found that a balanced droplet would suddenly move up or down, indicating that it had picked up an extra ion from the air.



110 Chapter 3 • Atoms

Millikan quickly learned to change the charge of the droplet at will by ionizing the air (by passing X-rays through the apparatus, for example). In this way he could measure not only the initial charge on the droplet but also the *change* in the charge as it acquired extra positive and negative ions.

In the course of several years, starting in 1906, Millikan observed thousands of droplets, sometimes watching a single droplet for several hours as it changed its charge a score or more times. He used several different liquids for his droplets and various procedures for changing the droplets' charges. In every case the original charge q and all subsequent changes in q were found to be integer multiples (positive and negative) of a single basic charge,

$$e = 1.6 \times 10^{-19} \text{ C}$$

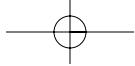
That the basic unit of charge was the same as the electron's charge was checked by pumping most of the air out of the apparatus. With few air molecules, there could be few ions for the droplet to pick up. Nevertheless, Millikan found that X-rays could still change the charge on the droplets, but only in the direction of increasing positive charge. He interpreted this, correctly, to mean that X-rays were knocking electrons out of the droplet. Since the changes in charge were still multiples of e (including sometimes e itself), it was clear that the electron's charge was itself equal to the unit of charge e (or rather $-e$, to be precise).

It is worth emphasizing the double importance of Millikan's beautiful experiments. First, he had measured the charge of the electron, $q = -e$. Combined with the measurement of m/e by Thomson and others, this also determined the mass of the electron (as about 1/2000 of the mass of the hydrogen atom, as Thomson had guessed). Second, and possibly even more important, he had established that *all* charges, positive and negative, come in multiples of the same basic unit e .

3.12 Rutherford and the Nuclear Atom*

As soon as the electron had been identified by Thomson, physicists began considering its role in the structure of atoms. In particular, Kelvin and Thomson developed a model of the atom (usually called the Thomson model), in which the electrons were supposed to be embedded in a uniform sphere of positive charge, somewhat like the berries in a blueberry muffin. The nature of the positive charge was unknown, but the small mass of the electron suggested that whatever carried the positive charge must account for most of the atom's mass.

To investigate these ideas, it was necessary to find an experimental probe of the atom, and the energetic charged particles ejected by radioactive substances proved suitable for this purpose. Especially suitable was the alpha particle, which is emitted by many radioactive substances and which Rutherford had identified (1909) as a positively ionized helium atom. (In fact, it is a helium atom that has lost both of its electrons; that is, it is a helium nucleus.) It was found that if alpha particles were directed at a thin layer of matter, such as a metal foil, the great majority passed almost straight through, suffering only small deflections. It seemed clear that the alpha particles must be passing through the atoms in their path and that the deflections must be caused by the electric fields inside the atoms. All of this was consistent with the Thomson model, which necessarily predicted only small deflections, since the electrons

Section 3.12 • Rutherford and the Nuclear Atom III

are too light and the field of the uniform positive charge too small to produce large changes in the velocity of a massive alpha particle.

However, Ernest Rutherford and his two assistants, Hans Geiger and Ernest Marsden, found (around 1910) that even with the thinnest of metal foils (of order $1\text{ }\mu\text{m}$) a few alpha particles were deflected through very large angles — 90° and even more. Since a single encounter with Thomson's atom could not possibly cause such a deflection, it was necessary to assume (if one wished to retain the Thomson model) that these large deflections were the result of many encounters, each causing a small deflection. But Rutherford was able to show that the probability of such multiple encounters was far too small to explain the observations.

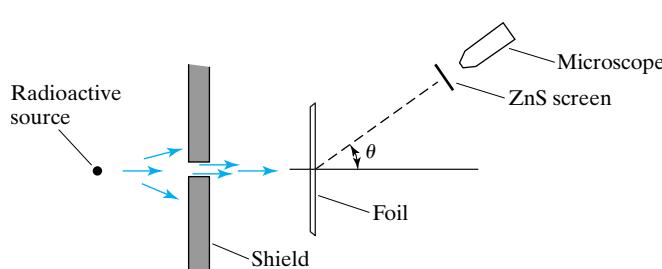
Rutherford argued that the atom must contain electric fields far greater than predicted by the Thomson model, and large enough to produce the observed large deflections in a single encounter. To account for these enormous electric fields he proposed his famous nuclear atom, with the positive charge concentrated in a tiny massive nucleus. (Initially, he did not exclude the possibility of a negative nucleus with positive charges outside, but it was quickly established that the nucleus was positive.)

According to Rutherford's model, the majority of alpha particles going through a thin foil would not pass close enough to any nuclei to be appreciably deflected. (Collisions with atomic electrons would not cause significant deflections because the electron is so much lighter than the alpha particles.) On the other hand, a few would come close to a nucleus, and these would be the ones scattered through a large angle. A simplifying feature of the model is that the large deflections occur close to the nucleus, well inside the electron orbits, and are unaffected by the electrons. That is, one can analyze the large deflections in terms of the Coulomb force of the nucleus, ignoring all the atomic electrons.

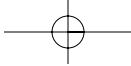
Rutherford was able to calculate the trajectory of an alpha particle in the Coulomb field of a nucleus and hence to predict the number of deflections through different angles; he also predicted how this number should vary with the particle's energy, the foil's thickness, and other variables. These predictions were published in 1911, and all were subsequently verified by Geiger and Marsden (1913), whose results provided the most convincing support for the nuclear atom.

The experiment of Geiger and Marsden is shown schematically in Fig. 3.8. A narrow stream of alpha particles from a radioactive source was directed toward a thin metal foil. The number of particles deflected through an angle θ was observed with the help of a zinc sulfide screen, which gave off a visible scintillation when hit by an alpha particle; these scintillations could be observed through a microscope and counted.

To calculate the number of deflections expected on the basis of Rutherford's model of the atom, let us consider a single alpha particle of mass m , charge $q = 2e$, and energy E . Since we are interested in large deflections

**FIGURE 3.8**

The “Rutherford scattering” experiment of Geiger and Marsden. Alpha particles from a radioactive source pass through a narrow opening in a thick metal shield and impinge on the thin foil. The number scattered through the angle θ is counted by observing the scintillations they cause on the zinc sulfide screen.



112 Chapter 3 • Atoms

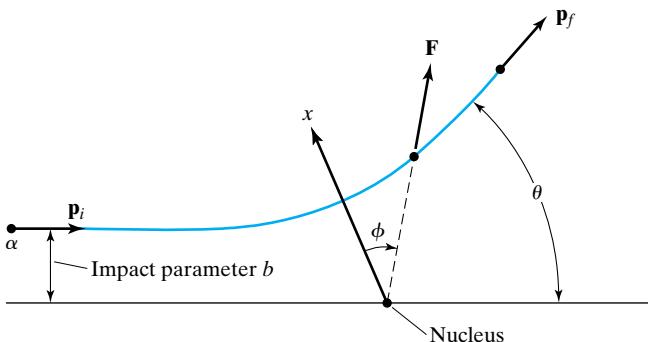


FIGURE 3.9

Trajectory of the alpha particle in the Coulomb field of a nucleus. The initial and final momenta are labeled \mathbf{p}_i and \mathbf{p}_f ; the direction labeled x is used as x axis in Section 3.13.

(θ more than a couple of degrees, say) we suppose that the particle passes reasonably close to the nucleus of an atom and well inside the atomic electrons, whose presence we can therefore ignore. We denote the nuclear charge by $Q = Ze$ and, for simplicity, suppose the nucleus to be fixed. (This is a good approximation since most nuclei are much heavier than the alpha particle.) The only force acting on the alpha particle is the Coulomb repulsion of the nucleus,

$$F = \frac{kqQ}{r^2} = \frac{2Zke^2}{r^2} \quad (3.45)$$

where k is the Coulomb force constant* $k = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$. Under the influence of this inverse-square force the alpha particle follows a hyperbolic path as shown in Fig. 3.9.

We can characterize the path followed by any particular alpha particle by its *impact parameter* b , defined as the perpendicular distance from the nucleus to the alpha particle's original line of approach (Fig. 3.9). Our first task is to relate the angle of deflection θ to the impact parameter b . We defer this somewhat tedious exercise in mechanics to Section 3.13, where we will prove that

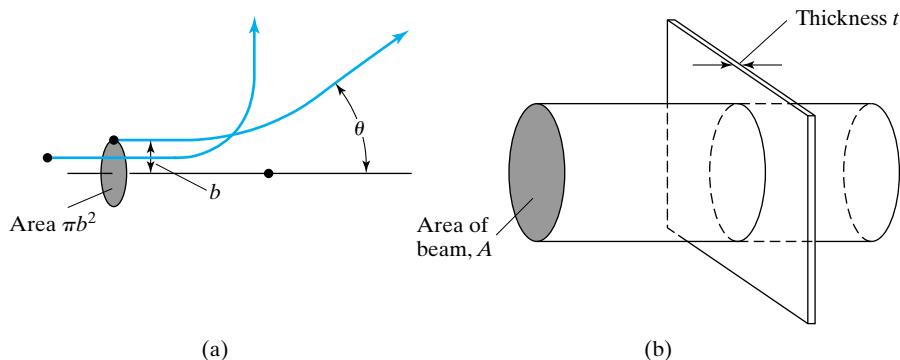
$$b = \frac{Zke^2}{E \tan(\theta/2)} \quad (3.46)$$

where E denotes the energy of the incident alpha particles. Notice that a large scattering angle θ corresponds to a small impact parameter b , and vice versa, just as one would expect.

Let us now focus on a particular value of the impact parameter b and the corresponding angle θ . All particles whose impact parameter is less than b will be deflected by more than θ [Fig. 3.10(a)]. Thus all those particles that impinge on a circle of radius b (and area πb^2) are scattered by θ or more. If the original beam of particles has cross-sectional area A [Fig. 3.10(b)], the proportion of particles scattered by θ or more is $\pi b^2/A$. If the total number of particles is N , the number scattered through θ or more by any one atom in the foil is

$$\text{(number scattered through } \theta \text{ or more by one atom)} = N \frac{\pi b^2}{A} \quad (3.47)$$

*The Coulomb force constant k is often written in the form $k = 1/(4\pi\varepsilon_0)$, where ε_0 is called the permittivity of the vacuum.

**FIGURE 3.10**

(a) A particle with impact parameter b is deflected by an angle θ ; all those particles that impinge on the circle of area πb^2 are deflected by more than θ . (b) The cross-sectional area of the whole beam is A ; the volume of target intersected by the beam is At .

This must now be multiplied by the number of target atoms that the beam of alpha particles encounters. If the target foil has thickness t and contains n atoms in unit volume, the number of atoms that the beam meets is [Fig. 3.10(b)]

$$(\text{number of target atoms encountered}) = nAt \quad (3.48)$$

Combining (3.47) and (3.48), we find that the total number of alpha particles scattered through θ or more is

$$N_{\text{sc}}(\theta \text{ or more}) = \frac{N\pi b^2}{A} \times nAt = \pi Nntb^2 \quad (3.49)$$

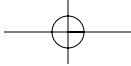
By differentiating (3.49) with respect to θ , we can find the number of particles emerging between θ and $\theta + d\theta$; and finally, by elementary geometry we can find the number that hit unit area on the zinc sulfide screen at angle θ and distance s from the foil. We again defer the details of the calculation to Section 3.13. The result is that

$$\begin{aligned} n_{\text{sc}}(\theta) &= \text{number of particles per unit area at } \theta \\ &= \frac{Nnt}{4s^2} \cdot \left(\frac{Zke^2}{E} \right)^2 \cdot \frac{1}{\sin^4(\theta/2)} \end{aligned} \quad (3.50)$$

This important result is called the **Rutherford formula**.

Some features of the Rutherford formula would be expected of almost any reasonable atomic model. For example, it is almost inevitable that $n_{\text{sc}}(\theta)$ should be proportional to N , the original number of alpha particles, and inversely proportional to s^2 , the square of the distance to the detector. On the other hand, several features are specific to Rutherford's assumption that each large angle deflection results from an encounter with the tiny but massive charged nucleus of a single target atom. Among the features specific to Rutherford's model are

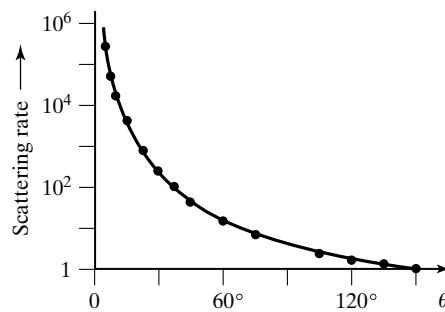
1. $n_{\text{sc}}(\theta)$ is proportional to the thickness t of the target foil.
2. $n_{\text{sc}}(\theta)$ is proportional to Z^2 , the nuclear charge squared.
3. $n_{\text{sc}}(\theta)$ is inversely proportional to E^2 , the incident energy squared.
4. $n_{\text{sc}}(\theta)$ is inversely proportional to the fourth power of $\sin(\theta/2)$.



114 Chapter 3 • Atoms

FIGURE 3.11

Geiger and Marsden measured the flux of scattered alpha particles at 14 different angles. Their measurements fit Rutherford's predicted $1/\sin^4(\theta/2)$ behavior beautifully. Notice the vertical scale is logarithmic and the measurements span an amazing 5 orders of magnitude.



Geiger and Marsden were able to check each of these predictions separately and found excellent agreement in all cases. Their observations of item (4), the variation the scattering with the angle θ , are plotted in Fig. 3.11, which shows their measurements taken at 14 different angles* and fitted to a curve of the form $K/\sin^4(\theta/2)$. These and their other equally beautiful results established Rutherford's nuclear atom beyond reasonable doubt and paved the way for the modern quantum theory of the atom.

One final consequence of Rutherford's analysis deserves mention. The Rutherford formula (3.50) was derived by assuming that the force of the nucleus on the alpha particles is the Coulomb force

$$F = \frac{kqQ}{r^2} \quad (3.51)$$

This is true provided that the alpha particles remain outside the nucleus at all times. If the alpha particles penetrated the nucleus, the force would not be given by (3.51) and the Rutherford formula would presumably not hold. Thus the fact that Geiger and Marsden found the Rutherford formula to be correct in all cases indicated that all alpha particles were deflected before they could penetrate the target nuclei. Since it was easy to calculate the minimum distance of the alpha particle from the center of any nucleus, this gave an upper limit on the nuclear radius, as we see in the following example.

Example 3.5

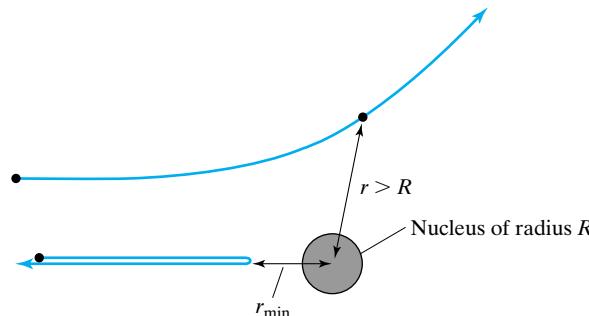
When 7.7-MeV alpha particles are fired at a gold foil ($Z = 79$), the Rutherford formula agrees with the observations at all angles. Use this fact to obtain an upper limit on the radius R of the gold nucleus.

Since the Rutherford formula holds at all angles, none of the alpha particles penetrate inside the nucleus; that is, for all trajectories the distance r from the alpha particle to the nuclear center is always greater than R

$$r > R \quad (3.52)$$

(See Fig. 3.12.) Now, the minimum value of r occurs for the case of a head-on collision, in which the alpha particle comes instantaneously to rest. At this

*Data is taken from Geiger and Marsden, *Philosophical Magazine*, vol. 25, pp. 604–623, (1913) — a surprisingly readable paper.

**FIGURE 3.12**

If none of the alpha particles penetrates into the nucleus, $r > R$ for all points on all orbits. The closest approach r_{\min} occurs in a head-on collision.

point its kinetic energy is zero ($K = 0$) and its potential energy $U = 2Zke^2/r_{\min}$, is equal to its total energy; that is, $U = E = 7.7 \text{ MeV}$. Thus

$$\frac{2Zke^2}{r_{\min}} = E \quad (3.53)$$

whence

$$r_{\min} = \frac{2Zke^2}{E} \quad (3.54)$$

Since $R < r$ for all orbits, it follows that $R < r_{\min}$ and hence

$$R \lesssim \frac{2Zke^2}{E} \quad (3.55)$$

We have used the sign \lesssim to emphasize that this is only an approximate result. In the first place, the nuclear radius is itself only an approximate notion. Second, we are ignoring the size of the alpha particle. [We could improve on (3.55) by replacing R by $R + R_\alpha$.] Finally, it is possible that non-Coulombic, nuclear forces may have an appreciable effect even a little before the alpha particle penetrates the nucleus.

Before we substitute numbers into the inequality (3.55), it is convenient to note that the Coulomb constant, $k = 9 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$, almost always appears in atomic and nuclear calculations in the combination ke^2 . Since ke^2/r is an energy, ke^2 has the dimensions of energy \times length and is conveniently expressed in the units $\text{eV} \cdot \text{nm}$ or, what is the same thing, $\text{MeV} \cdot \text{fm}$. To this end, we detach one factor of e and write

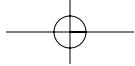
$$\begin{aligned} ke^2 &= (8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times (1.60 \times 10^{-19} \text{ C}) \times e \\ &= 1.44 \times 10^{-9} \text{ e} \cdot \text{N} \cdot \text{m}^2/\text{C} \end{aligned}$$

Now the unit $\text{N} \cdot \text{m}/\text{C} = \text{J}/\text{C}$ is the same as a volt. Multiplied by e , this gives an electron volt; therefore, $ke^2 = 1.44 \times 10^{-9} \text{ eV}$ or

$$ke^2 = 1.44 \text{ eV} \cdot \text{nm} = 1.44 \text{ MeV} \cdot \text{fm} \quad (3.56)$$

Returning to the inequality (3.55), we find that

$$R \lesssim \frac{2Zke^2}{E} = \frac{2 \times 79 \times (1.44 \text{ MeV} \cdot \text{fm})}{7.7 \text{ MeV}}$$



116 Chapter 3 • Atoms

or

$$R \lesssim 30 \text{ fm} \quad (3.57)$$

That is, the fact that Geiger and Marsden found the Rutherford formula valid for 7.7 MeV alpha particles on gold implied that the gold nucleus has radius less than 30 fm. Since the radius of the gold nucleus is in fact about 8 fm, this was perfectly correct.

If the incident energy E were steadily increased, some of the alpha particles would eventually penetrate the target nucleus, and the Rutherford formula would cease to hold, first at $\theta = 180^\circ$ (that is, for head-on collisions), then as the energy increased, at smaller angles as well. From (3.55) we see that the energy at which the Rutherford formula first breaks down is

$$E \approx \frac{2Zke^2}{R} \quad (3.58)$$

For gold this energy is about 30 MeV, an energy that was not available to Rutherford until much later. However, for aluminum, with Z only 13, it is about 6 MeV (Problem 3.49), and Rutherford was able to detect some departure from his formula at large angles.

3.13 Derivation of Rutherford's Formula*

In deriving the Rutherford formula (3.50), we omitted two slightly tedious calculations. First, the relation (3.46) between impact parameter b and scattering angle θ : If we call the initial and final momenta of the alpha particle \mathbf{p}_i and \mathbf{p}_f , the angle between \mathbf{p}_i and \mathbf{p}_f is the scattering angle θ , as shown in Fig. 3.13. Since the alpha particle's energy is unchanged, \mathbf{p}_i and \mathbf{p}_f are equal in magnitude and the triangle on the left of Fig. 3.13 is isosceles. If we denote the change in momentum by $\Delta\mathbf{p} = \mathbf{p}_f - \mathbf{p}_i$, it is clear from the construction shown that

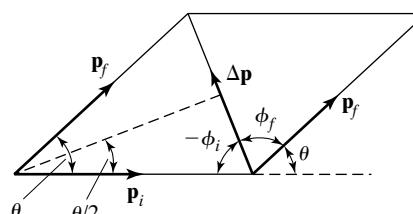
$$\Delta p = 2p_i \sin \frac{\theta}{2} \quad (3.59)$$

Now, we know from Newton's second law ($d\mathbf{p}/dt = \mathbf{F}$) that

$$\Delta\mathbf{p} = \int_{-\infty}^{\infty} \mathbf{F} dt$$

FIGURE 3.13

Geometry of the initial and final momenta in Rutherford scattering.





Section 3.13 • Derivation of Rutherford's Formula 117

To evaluate this integral, we choose our x axis in the direction of $\Delta\mathbf{p}$. This direction was indicated in Fig. 3.9, where we labeled the polar angle of the alpha particle as ϕ . With these notations

$$\Delta p = \int_{-\infty}^{\infty} F_x dt = \int_{-\infty}^{\infty} \frac{2Zke^2}{r^2} \cos \phi dt \quad (3.60)$$

This integral can be easily evaluated by a trick that exploits conservation of angular momentum. (Since the force on the alpha particle is radially outward from the nucleus, the angular momentum L about the nucleus is constant.) Long before the collision $L = bp_i$, whereas we know that at all times

$$L = mr^2\omega = mr^2 \frac{d\phi}{dt}$$

Equating these two expressions for L , we find that

$$\frac{d\phi}{dt} = \frac{bp_i}{mr^2}$$

Thus we can rewrite (3.60) as

$$\begin{aligned} \Delta p &= 2Zke^2 \int \frac{\cos \phi}{r^2} \cdot \frac{dt}{d\phi} d\phi \\ &= 2Zke^2 \int \frac{\cos \phi}{r^2} \cdot \frac{mr^2}{bp_i} d\phi \\ &= \frac{2Zke^2 m}{bp_i} \int_{\phi_i}^{\phi_f} \cos \phi d\phi \\ &= \frac{2Zke^2 m}{bp_i} (\sin \phi_f - \sin \phi_i) \end{aligned} \quad (3.61)$$

Comparing Figs. 3.9 and 3.13, we see that $\phi_f = -\phi_i$ and that $\theta + 2\phi_f = 180^\circ$. Therefore, $\phi_f = 90^\circ - \theta/2$ and $\sin \phi_f = \cos(\theta/2)$. Thus (3.61) becomes

$$\Delta p = \frac{4Zke^2 m}{bp_i} \cos \frac{\theta}{2} \quad (3.62)$$

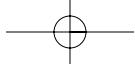
Equating the two expressions (3.59) and (3.61) for Δp , we see that

$$2p_i \sin \frac{\theta}{2} = \frac{4Zke^2 m}{bp_i} \cos \frac{\theta}{2}$$

which we can solve for b :

$$b = \frac{Zke^2}{E \tan(\theta/2)} \quad (3.63)$$

since $E = p_i^2/2m$. This completes the proof of the relation (3.46).



118 Chapter 3 • Atoms

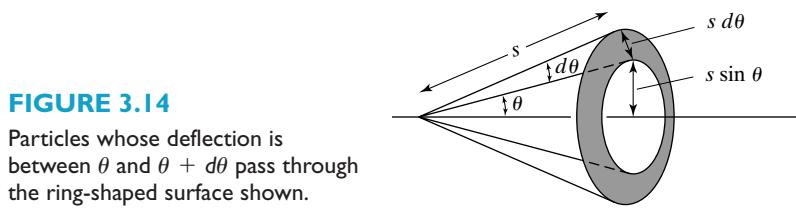


FIGURE 3.14

Particles whose deflection is between θ and $\theta + d\theta$ pass through the ring-shaped surface shown.

The other step that was omitted in Section 3.12 was the derivation of the final result, the Rutherford formula (3.50), from Equation (3.49),

$$N_{sc}(\theta \text{ or more}) = \pi N n t b^2$$

Substituting (3.63), we can rewrite the latter as

$$N_{sc}(\theta \text{ or more}) = \pi N n t \left(\frac{Z k e^2}{E \tan(\theta/2)} \right)^2 \quad (3.64)$$

This gives the number of alpha particles deflected through θ or more. The number that emerge between θ and $\theta + d\theta$ is found by differentiating (3.64) to give (see Problem 3.51)

$$N_{sc}(\theta \text{ to } \theta + d\theta) = \pi N n t \left(\frac{Z k e^2}{E} \right)^2 \frac{\cos(\theta/2)}{\sin^3(\theta/2)} d\theta. \quad (3.65)$$

Now, at a distance s from the target (where the detector is placed) the particles emerging between θ and $\theta + d\theta$ are distributed uniformly over the ring-shaped surface shown in Fig. 3.14. This ring has area

$$\text{area of ring} = (2\pi s \sin \theta) \times (s d\theta) \quad (3.66)$$

To find the number of particles per unit area at distance s , we must divide the number (3.65) by this area, to give

$$\begin{aligned} n_{sc}(\theta) &= \frac{N_{sc}(\theta \text{ to } \theta + d\theta)}{2\pi s^2 \sin \theta d\theta} \\ &= \frac{N n t}{4s^2} \cdot \left(\frac{Z k e^2}{E} \right)^2 \cdot \frac{1}{\sin^4(\theta/2)} \end{aligned}$$

where we have used the identity $\sin(\theta) = 2 \sin(\theta/2) \cos(\theta/2)$. This completes our derivation of the Rutherford formula.

CHECKLIST FOR CHAPTER 3

CONCEPT	DETAILS
Elements and compounds, atoms and molecules	Sec. 3.2
Electrons, nuclei, protons, and neutrons	Building blocks of atoms (Sec. 3.3)
Rutherford scattering	Scattering of alpha particles off a metal foil; first evidence for the nuclear atom (Secs. 3.3, 3.12 and 3.13)

Nucleon	Name for a proton or neutron
Unit of charge e	$e = -($ electron charge) = (proton charge)
Atomic radii	$\approx 0.1 \text{ nm} = 10^{-10} \text{ m}$ (3.1)
Nuclear radii	$\approx \text{a few fm} = \text{a few} \times 10^{-15} \text{ m}$ (3.3)
Atomic number of an atom Z	$Z = (\text{number of electrons in neutral atom})$ $= (\text{number of protons in nucleus})$ (3.4)
Ions	Atoms that have lost or gained one or more electrons
Relative masses of e, p, and n	$m_p \approx m_n \approx 2000 m_e$ (3.5)
Mass number A	$A = \text{number of nucleons (protons + neutrons) in nucleus}$ (3.7)
Isotopes	Two or more atoms with the same atomic number Z but different neutron number. Same chemical properties, but different masses
Atomic mass unit, u	$1 \text{ u} = \frac{1}{12}(\text{mass of neutral atom of } ^{12}\text{C}) \approx (\text{mass of } ^1\text{H atom})$ (3.12)
Avogadro's constant, N_A	$N_A = 6.022 \times 10^{23} \text{ objects/mole}$ (3.15)
Mole (mol)	1 mole = N_A objects
Ideal gas law	$pV = nRT = Nk_B T$ (3.17) and (3.18) derivation from kinetic theory (Sec. 3.7)
Universal gas constant, R	$R = 8.31 \text{ J/K} \cdot \text{mole}$
Boltzmann's constant, k_B	$k_B = R/N = 1.38 \times 10^{-23} \text{ J/K} = 8.62 \times 10^{-5} \text{ eV/K}$ (3.19)
Mean free path, λ	Mean distance between collisions (Sec. 3.8)
Diffusive motion	Random-walk motion. Displacement $D_{\text{dif}} \propto \sqrt{t}$ (Sec. 3.8)
Brownian motion	Jiggling of small particles due to impacts of atoms (Sec. 3.9)
Thomson's discovery of the electron*	Sec. 3.10
Millikan's oil-drop experiment*	Measurement of electron charge (Sec. 3.11)
Rutherford's scattering experiment*	Established nuclear model of atom (Secs. 3.12 and 3.13)

PROBLEMS FOR CHAPTER 3

SECTION 3.2 (Elements, Atoms, and Molecules)

- 3.1** • One of the triumphs of nineteenth-century science was the discovery by Mendeleev of the periodic table of the elements, which brought order to the confusing multitude of known elements, as described in Chapter 10. A crucial step in this discovery was to list the elements in the correct order — that is, as we now know, in order of increasing atomic number (which we define in Section 3.4, as the number of electrons in the neutral atom). At the time, atomic number was an unknown concept, and chemists listed the elements in

order of atomic mass. Fortunately, these two orderings are very nearly the same. Use the periodic table inside the back cover of the book (where both parameters are listed) to find out how many times the two orders are different. The exception involving argon and potassium was a stumbling block in the development of the periodic table since it was found necessary to reverse the order of these two (when ordered by mass) to get the periodic table to make sense. We will see in Chapter 16 why it is that atomic number and atomic mass increase almost perfectly in step.

120 Chapter 3 • Atoms

3.2 • The elements are generally listed in order of their “atomic number” Z (which is just the number of electrons in a neutral atom of the element). Hydrogen is first with $Z = 1$, helium next with $Z = 2$, and so on. There are 90 elements that occur naturally on earth, but they are not just numbers 1 through 90. What are they? What are the atomic numbers of the known artificial elements? (You can find this information in the periodic table inside the back cover.)

3.3 • With what mass of oxygen would 2 g of hydrogen combine to form water (H_2O)?

[*HINT:* All you need to know is the ratio of the masses of the atoms concerned; you can find these masses, listed in atomic mass units, inside the back cover.]

3.4 • With what mass of carbon would 1 g of hydrogen combine to form acetylene, C_2H_2 ? (Read the hint to Problem 3.3.)

3.5 • With what mass of nitrogen would 1 g of hydrogen combine to form ammonia (NH_3)? (Read the hint to Problem 3.3.)

3.6 • (a) With what mass of oxygen would 1 g of carbon combine to form carbon monoxide CO ? (b) With what mass of oxygen would 1 g of carbon combine to form carbon dioxide CO_2 ? (In practice, when oxygen and carbon combine — as in a car engine — both CO and CO_2 are formed in proportions that depend on conditions such as the temperature.)

SECTION 3.4 (Some Atomic Parameters)

3.7 • Make a table showing the numbers of electrons, protons, neutrons, and nucleons in the most common form of the following atoms: He, Li, Na, Ca. (The necessary information is in Appendix D.)

3.8 • Tabulate the numbers of electrons, protons, neutrons, and nucleons in the most common form of the following neutral atoms: hydrogen, carbon, nitrogen, oxygen, aluminum, iron, lead. (The necessary information can be found in Appendix D.) On the basis of this information, can you suggest why lead is much denser than aluminum?

3.9 • Make a table similar to Table 3.5 showing all the stable isotopes for hydrogen, helium, oxygen, and aluminum, inclu their percentage abundances. (Use the information in Appendix D.)

3.10 • Use Appendix D to find four elements that have only one type of stable atom (that is, an atom with no stable isotopes).

3.11 • Two atoms that have the same mass number A but different atomic numbers Z are called *isobars*. (Isobar means “of equal mass.”) For example, ^3He and ^3H are isobars. Use the data in Appendix D to find three examples of pairs of isobars. Find an example of a triplet of isobars (that is, three different atoms with the same mass number).

3.12 • It is found that the radius R of any nucleus is given approximately by

$$R = R_0 A^{1/3}$$

where A is the mass number of the nucleus and R_0 is a constant whose value depends a little on how R is defined, but is about 1.1 fm. (a) What are the radii of the nuclei of helium, carbon, iron, lead, and lawrencium? (b) How does the volume of a nucleus (assumed spherical) depend on A ? What does your answer tell you about the average density of nuclei?

SECTION 3.5 (The Atomic Mass Unit)

3.13 • The mass of a carbon 12 atom is 1.992648×10^{-26} kg (with an uncertainty of 1 in the final digit). Calculate the value of the atomic mass unit to six significant figures in kg and in MeV/c^2 .

3.14 • Find the mass of a ^{12}C nucleus in u correct to five significant figures. (The mass of an electron is 0.000549 u. Ignore the binding energy of the electrons.)

3.15 • Using the natural abundances listed in Table 3.5 (or Appendix D) calculate the average atomic masses of naturally occurring carbon, chlorine, and iron. (Your answers should be in u and correct to three significant figures.)

3.16 • What is the mass (correct to the nearest u) of the following molecules: (a) water, H_2O ; (b) laughing gas, N_2O ; (c) ozone, O_3 ; (d) glucose, $\text{C}_6\text{H}_{12}\text{O}_6$; (e) ammonia, NH_3 ; (f) limestone, CaCO_3 .

3.17 • For the great majority of atoms, it is found that the atomic mass (in atomic mass units) is close to twice the atomic number (an observation we will explain in Chapter 16). (a) Using the data in the periodic table inside the back cover, verify this observation by computing the ratio

$$(\text{atomic mass in u})/(\text{atomic number})$$

for the elements with $Z = 10, 20, \dots, 90$. Comment on your results. (b) In terms of the mass number A , your results should show that A is close to (or in some cases a bit more than) $2Z$. What does this mean for the neutron number N in relation to the proton number Z ?

3.18 • (a) Calculate the mass of the ^4He *nucleus* in atomic mass units by subtracting the mass of two electrons from that of a neutral ^4He atom. (See Appendix D, and give four decimal places.) (b) The procedure suggested for part (a) is theoretically incorrect because we have neglected the binding energy of the electrons. Given that the energy needed to pull both electrons far away from the nucleus is 80 eV, is the correct answer larger or smaller than implied by part (a), and by how much (in u)? In stating the mass of the nucleus, about how many significant figures can you give before this effect would show up?

SECTION 3.6 (Avogadro's Number and the Mole)

- 3.19** • How many moles are in (a) 1 g of carbon? (b) 1 g of hydrogen molecules? (c) 10 g of water? (d) 1 ounce of gold?
- 3.20** • How many moles are in (a) 10 g of NaCl? (b) 2 kg of NH₃? (c) 10 cm³ of Hg (density 13.6 g/cm³)? (d) 1 pound of sugar (C₁₂H₂₂O₁₁)?
- 3.21** • A mole of O₂ molecules has mass 32 g (that is, the molecular “weight” of O₂ is 32 g/mole). What is the mass of a single O₂ molecule (in u and in grams)?
- 3.22** • What is the total charge of 1 mole of Cl⁻ ions? (The Cl⁻ ion is a chlorine atom that has acquired one extra electron.) This important quantity is called the faraday; it is easily measured since it is the charge that must pass to release 1 mole of chlorine in electrolysis of NaCl solution, (or, more generally, 1 mole of any monovalent ion in electrolysis of an appropriate solution). Once the electron charge was known, knowledge of the faraday let one calculate Avogadro’s number.
- 3.23** • (a) How many atomic mass units are there in one pound mass? (b) How many in one gram? Comment on your answer to part (b).
- 3.24** •• How many C atoms are there in (a) 1 g of CO₂? (b) 1 mole CH₄? (c) 1 kmole of C₂H₆? In each case what fraction of the atoms are C atoms, and what fraction of the mass is C?

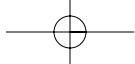
SECTION 3.7 Kinetic Theory

- 3.25** • Show that the number density (number per volume) of molecules in an ideal gas is p/kT .
- 3.26** • What is the edge length of a cube that contains 1 billion (10^9) molecules of a gas at standard temperature and pressure (STP: $T = 273\text{ K}$, $p = 1\text{ atm} = 1.01 \times 10^5\text{ N/m}^2$).
- 3.27** • At what pressure would the density of He gas at $T = 300\text{ K}$ equal the density of ordinary water? Give your answer in atmospheres. [$1\text{ atm} = 1.01 \times 10^5\text{ N/m}^2$].
- 3.28** •• Consider an ideal gas at standard temperature and pressure (STP: $T = 273\text{ K}$, $p = 1\text{ atm} = 1.01 \times 10^5\text{ N/m}^2$). Imagine that the molecules of the gas are equally spaced so that every molecule is at the center of a small cube. (a) What is the edge length of this cube, which is the average distance between nearest-neighbor molecules in the gas? (b) Qualitatively, how does this mean molecular separation distances compare with the diameter of a small molecule such as O₂?
- 3.29** ••• (a) Consider an ideal gas at temperature T with molecules of mass m and number density (number per volume) n . Show that the number of molecules that strike a surface of area A in a time Δt is given roughly by $(nA \Delta t/2)\sqrt{kT/m}$. To simplify this calculation, consider a wall perpendicular to the x -axis and make

the radical assumption that the x -components of the velocities of all molecules have the *same* magnitude; that is, assume that all molecules have the same $|v_x|$, with half the molecules moving to the right and the other half moving to the left. (b) For oxygen at standard temperature and pressure (STP: $T = 273\text{ K}$, $p = 1\text{ atm} = 1.01 \times 10^5\text{ N/m}^2$), what is the average number of particles that strike a 1 cm² surface in 1 s?

SECTION 3.8 The Mean Free Path and Diffusion

- 3.30** • Consider a large container of oxygen gas at STP ($T = 273\text{ K}$, $p = 1\text{ atm} = 1.01 \times 10^5\text{ N/m}^2$). Assume a molecular radius of 0.15 nm. (a) How long does it take for a molecule to diffuse to a point 1 m from its starting point? (b) How long would it take for the molecule to diffuse the same distance at a pressure of 0.01 atm? (c) Solve for the pressure at which a molecule will diffuse a displacement of magnitude d in a time t .
- 3.31** • (a) Show that for a gas, the mean free path λ between collisions is related to the mean distance between nearest neighbors r by the approximate relation $\lambda \approx r(r^2/\sigma)$ where σ is the collision cross-section. (b) Given that the molecular radius of a gas molecule such as O₂, N₂, or CO₂ is about 0.15 nm, estimate the value of r and λ for air at STP (standard temperature and pressure, $T = 273\text{ K}$, $p = 1.00\text{ atm} = 1.01 \times 10^5\text{ N/m}^2$). (c) Why, qualitatively, is $\lambda \gg r$, for dilute gases?
- 3.32** • The **Knudsen regime** is the pressure regime in which the mean free path is comparable to the dimensions of the container. (a) Given the container diameter d , the molecule’s collision cross section σ , and the temperature T , derive an expression for the pressure p at which the Knudsen regime occurs. (b) What is this pressure (in atmospheres) for a gas at room temperature (about 293 K) in a container with diameter 20 cm? Assume a molecular radius of 0.15 nm. A good mechanical pump can achieve a pressure of 10^{-4} atm; can such a pump produce Knudsen pressures in a 20-cm diameter container?
- 3.33** •• The best vacuum that can be readily achieved in the lab is called “ultra-high vacuum” or UHV and is 10^{-10} torr (mm Hg) $\approx 10^{-13}$ atm. At that pressure and at room temperature ($T = 293\text{ K}$), (a) how many air molecules are present in 1 cm³, (b) what is the mean free path λ of the molecules? Assume a molecular radius of 0.15 nm.
- 3.34** •• An aimless physics student, wandering around on a flat plane, takes a 1-m step in a random direction each second. (a) After one year of continuous random walking, what is the student’s expected distance from his starting point? (b) If the student wandered in 3D space, rather than in a plane, but still took 1-m steps each second in random directions, would his expected distance from the origin be greater, less, or the same as before. Explain.



122 Chapter 3 • Atoms

- 3.35** •• A gambler has \$100 and makes a series of \$1 bets that he can correctly guess the outcome of a coin flip — heads or tails. (The coin and the game are honest, so for every coin flip, there is a 50% chance that the gambler will guess correctly.) The gambler will quit when he has either won an additional \$100 or has lost all his money. About how many times will the gambler bet before quitting?

[HINT: This situation is like a random walk in one dimension.]

- 3.36** ••• Derive the ideal gas law without the assumption of no intermolecule collisions. To simplify this calculation, make the radical assumption that the x -components of the velocities of all molecules have the *same* magnitude; that is, assume that all molecules have the same $|v_x|$, with half the molecules moving to the right and the other half moving to the left; make similar assumptions for y - and z -directions. Consider a gas of molecules with number density n in a cubical container, and consider only those molecules that are within a short distance Δx of one wall of the container where $\Delta x \ll \lambda$, the mean free path. At any instant, half of those molecules are heading toward the wall and will collide with the wall in an average time Δt , where $\Delta x/\Delta t = |v_x|$. Proceed to compute the average force on the the wall due to molecular collisions, and then deduce the ideal gas law.

SECTION 3.9 Brownian Motion

- 3.37** •• A Brownian particle is observed to diffuse $100\text{ }\mu\text{m}$ in 60 s , on average. (Diffusion distance means magnitude of the net displacement.) A second particle with twice the diameter and ten times the mass of the first is observed. How far, on average, will the second particle diffuse in 60 s at the same temperature?
- 3.38** •• A Brownian particle is observed to diffuse a distance of $10\text{ }\mu\text{m}$ in 20 s , on average. How far, on average, will the particle diffuse in 200 s ?

- 3.39** •• A good microscope can readily resolve objects that are $1\text{ }\mu\text{m}$ in diameter ($1\text{ }\mu\text{m}$ is about twice the wavelength of visible light). The unaided human eye can resolve objects no smaller than roughly $0.1\text{ mm} = 100\text{ }\mu\text{m}$. Suppose a $1\text{-}\mu\text{m}$ diameter Brownian particle is observed (with a microscope) to diffuse an average distance of 5 times its diameter ($5\text{ }\mu\text{m}$) in 20 s . Under the same conditions of temperature and air viscosity, another Brownian particle of $100\text{ }\mu\text{m}$ diameter is observed with the unaided eye. How long will this particle take to diffuse an average distance of 5 times its diameter? Why was Brownian motion not discovered until after the invention of the microscope?

SECTION 3.10 (Thomson's Discovery of the Electron)

- 3.40** •• In Thomson's experiment electrons travel with velocity v in the x direction. They enter a uniform electric field E , which points in the y -direction and has

total width l (Fig. 3.15). Find the time for an electron to cross the field and the y -component of its velocity when it leaves the field. Hence show that its velocity is deflected through an angle $\theta \approx eEl/(mv^2)$ (provided that θ is small). Assume that the electrons are nonrelativistic, as was the case for Thomson.

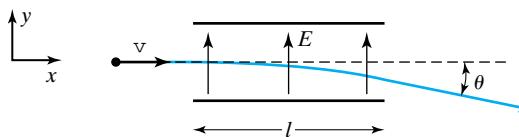


FIGURE 3.15

(Problem 3.40)

- 3.41** •• Suppose that the electrons in Thomson's experiment enter a uniform magnetic field B , which is in the z -direction (with axes defined as in Fig. 3.15) and has total width l . Show that they are deflected through an angle $\theta \approx eBl/(mv)$ (provided that θ is small). Assume that the electrons are nonrelativistic.

- 3.42** •• The formulas for the deflection of an electron in an electric or magnetic field (Problems 3.40 and 3.41) are quite similar. Explain in words why both formulas have factors of v in the denominator and why the magnetic deflection has just one power of v where the electric has two.

- 3.43** •• Thomson directed his cathode rays at a metal body and measured the total charge Q that it acquired and its rise in temperature ΔT . From ΔT and the body's known thermal capacity, he could find the heat given to the body. Show that this heat should be

$$\text{heat} = \frac{Qmv^2}{2e}$$

where m , v , and e are the electron's mass, speed (non-relativistic), and charge. Combine this with the result of Problem 3.41 to give expressions for v and m/e in terms of measured quantities.

- 3.44** •• In one version of Thomson's measurement of m/e , nonrelativistic electrons are accelerated through a measured potential difference V_0 . This means that their kinetic energy is $K = eV_0$. They are then passed into a known magnetic field B and the radius $R = mv/(eB)$ of their circular orbit is measured. Eliminate the speed v from these two equations, and derive an expression for m/e in terms of the measured quantities V_0 , B , and R .

SECTION 3.11 (Millikan's Oil-Drop Experiment)

- 3.45** ••• In order to find the electron's charge e , Millikan needed to know the mass M of his oil drops, and this was actually the source of his greatest uncertainty in determining e . (His value of e was about 0.5% low as a result.) However, to show that all charges are multiples of some basic unit charge, it was *not* necessary to

know M , which cancels out of the charge ratios. The following problem illustrates this point and gives some more details of Millikan's experiment.

By switching the E field off and on, Millikan could time an oil drop as it fell and rose through a measured distance l (falling under the influence of gravity, rising under that of E and gravity). The downward and upward speeds are given by (3.42) as

$$v_d = \frac{Mg}{6\pi r\eta} \quad \text{and} \quad v_u = \frac{qE - Mg}{6\pi r\eta} \quad (3.67)$$

Both speeds were measured in the form $v = l/t$, where t is the time for the droplet to traverse l .

- (a) By adding the two equations (3.67) and rearranging, show that

$$\frac{1}{t_d} + \frac{1}{t_u} = \left(\frac{E}{6\pi r\eta l} \right) q = Kq$$

where the quantity $K = E/(6\pi r\eta l)$ was constant as long as Millikan watched a single droplet and did not vary E .

From (3.68) we see that the charge q is proportional to the quantity $(1/t_d) + (1/t_u)$. Thus if it is true that q is always an integer multiple of e , it must also be true that $(1/t_d) + (1/t_u)$ is always an integer multiple of some fixed quantity. Table 3.7 shows a series of timings for a single droplet (made with a commercial version of the Millikan experiment used in a teaching laboratory). That t_u changes abruptly from time to time indicates that the charge on the droplet has changed as described in Section 3.11. (The times t_d should theoretically all be the same, of course. The small variations give you a good indication of the uncertainties in all timings. For t_d you should use the average of all measurements of t_d .)

Table 3.7

A series of measurements of t_d and t_u , the times for a single droplet to travel down and up a fixed distance l . All times are in seconds. (Problem 3.45)

$t_d:$	15.2	15.0	15.1	15.0	14.9	15.1	15.1	15.0	15.2	15.2
$t_u:$	6.4	6.3	6.1	24.4	24.2	3.7	3.6	1.8	2.0	1.9

- (b) Calculate $(1/t_d) + (1/t_u)$ and show that (within the uncertainties) this quantity is always an integer multiple of one fixed number, and hence that the charge is always a multiple of one fixed charge.
- (c) Use Equation (3.44) for v_d and the additional data in Table 3.8 to find the radius r of the droplet.
- (d) Finally, use (3.68) to find the four different charges q on the droplet. What is the best estimate of e based on this experiment?

Table 3.8

Additional data for the oil-drop experiment.

$l = 8.3 \times 10^{-4}$ m	(distance traveled by the droplet)
$\rho = 839$ kg/m ³	(density of oil)
$g = 9.80$ m/s ²	(acceleration of gravity)
$\eta = 1.60 \times 10^{-5}$ N·s/m ²	(viscosity of air, adjusted*)
$E = 1.21 \times 10^5$ N/C	(electric field)

*For very small droplets, there is a small correction to the formula (3.42) for the terminal velocity. For the experiment reported here, this correction amounts to a 12% reduction in the viscosity. We have included this correction in the value given.

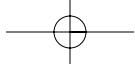
SECTION 3.12 (Rutherford and the Nuclear Atom)

3.46 • A student doing the Rutherford scattering experiment arranges matters so that she gets 80 counts/min at a scattering angle of $\theta = 10^\circ$. If she now moves her detector around to $\theta = 150^\circ$, keeping it at the same distance from the target, how many counts would she expect to observe in a minute? (This illustrates an awkward feature of the experiment, especially before the days of automatic counters. An arrangement that gives a reasonable counting rate at small θ gives far too few counts at large θ , and one that gives a reasonable rate for large θ will overwhelm the counter at small θ .)

3.47 •• The Rutherford model of the atom could explain the large-angle scattering of alpha particles because it led to very large electric fields compared to the Thomson model. To see this, note that in the Thomson model the positive charge of an atom was uniformly distributed through a sphere of the same size as the atom itself. According to this model, what would be the maximum E field (in volts/meter) produced by the positive charge in a gold atom ($Z = 79$, atomic radius ≈ 0.18 nm)? What is the corresponding maximum field in Rutherford's model of the gold atom, with the positive charge confined to a sphere of radius about 8 fm? (In the Thomson model the actual field would be even less because of the electrons. Note that since $ke^2 = 1.44$ eV·nm, it follows that $ke = 1.44$ V·nm.)

3.48 •• Consider several different alpha particles approaching a nucleus, with various different impact parameters, b , but all with the same energy, E . Prove that an alpha particle that approaches the nucleus head-on ($b = 0$) gets closer to the nucleus than any other.

3.49 •• (a) If the Rutherford formula is found to be correct at all angles when 15-MeV alpha particles are fired at a silver foil ($Z = 47$), what can you say about the radius of the silver nucleus? (b) Aluminum has atomic number $Z = 13$ and a nuclear radius about $R_{Al} \approx 4$ fm. If one were to bombard an aluminum foil with alpha particles and slowly increase their energy, at about what energy would you expect the Rutherford formula to break down? You can make the estimate (3.58) a bit more realistic by taking R to



124 Chapter 3 • Atoms

be $R_{\text{Al}} + R_{\text{He}}$, where R_{He} is the alpha particle's radius (about 2 fm).

- 3.50** •• In a student version of the Rutherford experiment ^{210}Po is used as a source of 5.2-MeV alpha particles, which are directed at a gold foil (thickness 2 μm) at a rate of 10^5 particles per minute. The scattered particles are detected on a screen of area 1 cm^2 at a distance of 12 cm. Use the Rutherford formula (3.50) to predict the number of alpha particles observed in 10 minutes at $\theta = 15^\circ, 30^\circ$, and 45° , given the following data:

$$\begin{aligned} N &= 10^6 && (\text{number of incident particles in 10 min}) \\ \rho &= 19.3 \text{ g/cm}^3 && (\text{density of gold}) \\ m_{\text{Au}} &= 197 \text{ u} && (\text{mass of gold atom}) \\ t &= 2 \times 10^{-6} \text{ m} && (\text{thickness of foil}) \end{aligned}$$

$$Z = 79 \quad (\text{atomic number of gold})$$

$$E = 5.2 \text{ MeV} \quad (\text{energy of alpha particles})$$

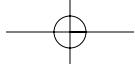
[Note that the number density n of gold nuclei is ρ/m_{Au} , and don't forget that $ke^2 = 1.44 \text{ MeV} \cdot \text{fm}$.]

SECTION 3.13 (Derivation of Rutherford's Formula)

- 3.51** • (a) If $N_{\text{sc}}(\theta \text{ or more})$ denotes the number of alpha particles scattered by θ or more, show that the number whose deflection is between θ and $\theta + d\theta$ is

$$N_{\text{sc}}(\theta \text{ to } \theta + d\theta) = -\frac{dN_{\text{sc}}(\theta \text{ or more})}{d\theta} d\theta$$

- (b) Differentiate the expression (3.64) for $N_{\text{sc}}(\theta \text{ or more})$ and verify the result (3.65).



Chapter 4

Quantization of Light

- 4.1 Quantization
 - 4.2 Planck and Blackbody Radiation
 - 4.3 The Photoelectric Effect
 - 4.4 X-rays and Bragg Diffraction
 - 4.5 X-ray Spectra
 - 4.6 The Compton Effect
 - 4.7 Particle-Wave Duality
- Problems for Chapter 4

4.1 Quantization

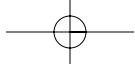
Quantum theory replaced classical physics in the microscopic domain because classical physics proved incapable of explaining a wide range of microscopic phenomena. In this chapter we describe the first clear evidence for this failure of classical physics, the evidence that light and all other forms of electromagnetic radiation are *quantized*.

We say that something is **quantized** if it can occur only in certain discrete amounts. Thus “quantized” is the opposite of “continuous.” (The word comes from the Latin “quantum,” meaning “how much.”) In our everyday experience at the store, eggs are quantized (since one can take only integer numbers of eggs), whereas gasoline is not (since one can take any amount of gasoline). On the microscopic level, we have seen that matter is quantized, the smallest unit, or *quantum*, of normal matter being the atom. Similarly, we have seen that electric charge is quantized, the quantum of charge being $e = 1.6 \times 10^{-19} \text{ C}$.

Early in the twentieth century it was found that electromagnetic radiation is quantized. The energy contained in a beam of light* of a given frequency is not a continuous variable; instead, it can only be an integer multiple of a certain basic quantum of energy. In fact, it was found that an electromagnetic wave consists of tiny localized bundles of energy, which also carry momentum and have most of the properties of ordinary particles, except that their mass is zero. These bundles, or quanta of light, have come to be called **photons**.

The quantization of matter and of electric charge had not contradicted any basic principles of classical physics. On the other hand, the quantization of radiation was definitely inconsistent with classical electromagnetic theory, which predicted unambiguously that the energy of radiation of any given frequency should be a continuous variable. Thus the discovery that light is quantized required the development of a new theory, the quantum theory of radiation.

*When there is no serious risk of confusion we use “light” to mean any form of electromagnetic radiation. When necessary, we use “visible light” to emphasize that we are speaking only of the radiation to which the human eye is sensitive.



126 Chapter 4 • Quantization of Light

4.2 Planck and Blackbody Radiation

Max Planck
(1858–1947, German)



Planck is best known for suggesting (in 1900) the quantization of radiant energy — an idea that won him the 1918 Nobel Prize in physics. He was quick to take up Einstein's relativity in 1905 and was the first to propose the correct relativistic expression for the momentum of a massive particle (in 1906). By 1930, his prestige was such that the Kaiser Wilhelm Society of Berlin was renamed the Max Planck Society, and Planck was made its president; but in 1937 he was forced to resign because of his opposition to the persecution of Jews in Nazi Germany.

The first person to propose that electromagnetic radiation must be quantized was the German physicist Max Planck, in connection with his studies of blackbody radiation in the year 1900. A blackbody is any body that is a perfect absorber of radiation, and blackbody radiation is the radiation given off by such a body when heated. (In practice, one can realize blackbody radiation by observing the radiation coming out of a small hole in an enclosed chamber or “oven.”) For some 40 years, physicists had been trying to calculate how the energy in blackbody radiation was distributed with respect to wavelength (how much energy at any given wavelength). Planck himself had favored a theory proposed by the German physicist Wilhelm Wien, but by 1900 experiments were showing that this theory could not be correct. In 1900 (just a few months before Planck’s quantum proposal) the English physicist Lord Rayleigh had proposed a formula based on classical electromagnetism and classical statistical mechanics. Rayleigh’s formula (now generally called the Rayleigh–Jeans formula, to honor James Jeans, who had discovered a small correction to Rayleigh’s original formula) fitted the data well at very long wavelengths, but was hopelessly wrong at shorter wavelengths, as you can see in Fig. 4.1, where the experimental energy distribution is shown by the dots and the Rayleigh–Jeans formula by the dashed curve.

Then, at a meeting of the German Physical Society in October of 1900, Planck announced that he had found a formula that seemed to fit all of the experimental data. He arrived at his formula by a fairly devious route, but he was eventually able to show that it hinged on the assumption that the radiation emitted by a body is quantized. Specifically, he assumed that radiation of frequency f can be emitted only in integral multiples of a basic quantum hf ,

$$E = 0, hf, 2hf, 3hf, \dots \quad (4.1)$$

where h was an unknown constant, now called **Planck’s constant**. Notice that Planck’s quantum of energy, hf , varies with the frequency. This contrasts with the quantum of charge e , which is the same for all charges.

Planck showed that the assumption (4.1) led to his new formula (the Planck formula) for blackbody radiation. This formula depended on the unknown constant h , which he chose so as to give the best fit to the experimental data. With h chosen in this way, Planck’s formula fits the data perfectly at all

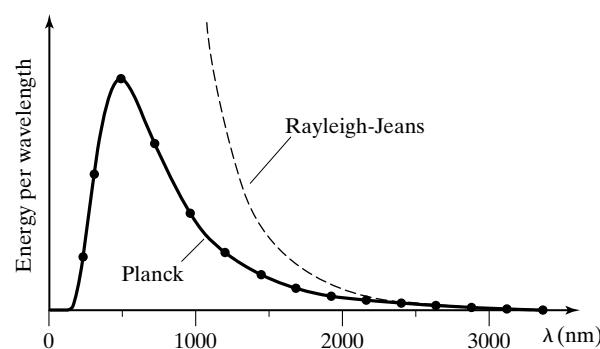


FIGURE 4.1

The distribution of intensity with wavelength of blackbody radiation at $T = 6000$ K. (This is the approximate temperature of the sun’s surface, which is quite nearly a perfect blackbody.) The classical Rayleigh–Jeans formula (dashed curve) agrees with the experimental points (shown as dots) at very long wavelengths but is hopelessly wrong at shorter wavelengths. Planck’s formula (solid curve), based on the quantization of light, agrees perfectly with experiment. Notice that, at this temperature, the intensity is greatest at about 500 nm, which explains why sunlight is white. It also explains why the human eye evolved with its sensitivity in the region around 500 nm.

frequencies and all temperatures, as illustrated in Fig. 4.1. The modern value of Planck's constant h is

$$h = 6.63 \times 10^{-34} \text{ J} \cdot \text{s} \quad (4.2)$$

Notice that since hf is an energy, h has the dimensions of energy \times time.

It would be hard to overstate the importance of Planck's work on blackbody radiation since it clearly marks the beginning of quantum theory, which revolutionized physics and chemistry and spawned so much of the technology that makes possible the lifestyle we enjoy today — transistors, radio, television, X-rays, magnetic resonance imaging, computers, and so much more. Nevertheless, the details of blackbody radiation are surprisingly unimportant in most of what we will discuss in the remainder of this book. Furthermore, few introductory physics courses cover the material (notably, statistical mechanics) needed to appreciate Planck's ideas. Therefore, we will not pursue them further here. If you would like some more details, please look at Problems 4.1–4.4 and 4.32–4.33 at the end of the chapter.* In particular, you can find Planck's formula in Eq. (4.28) of Problem 4.1. For now, the main thing to learn is that Planck's suggestion that the energy of electromagnetic radiation is quantized in accordance with (4.1) proved correct and was the starting point of modern quantum theory.

4.3 The Photoelectric Effect

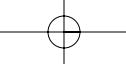
In 1905 Planck's ideas were taken up and extended by Einstein, who showed that they explain several phenomena, the most important of which was the **photoelectric effect**. In this effect, discovered by Heinrich Hertz in 1887, a metal exposed to light is found to eject electrons from its surface. At first sight this process, which is the basis of some modern light-detecting devices, appeared perfectly consistent with classical electromagnetic theory. Light waves were known to carry energy in the form of oscillating electric and magnetic fields, and it seemed perfectly reasonable that some electrons in the metal could absorb enough of this energy to be ejected. However, closer investigation showed that several features of the process were incompatible with classical electromagnetic theory.

An apparatus for investigating the photoelectric effect is sketched in Fig. 4.2. Light is shone on one of the two electrodes in an evacuated glass tube, and electrons are ejected. If the other electrode is kept at a higher potential, it attracts these electrons, causing a current, whose magnitude indicates the number of electrons being ejected. If, instead, the second electrode is kept at a lower potential, it repels the electrons and only those electrons with enough kinetic energy to overcome the retarding potential V reach the second electrode. As one increases the retarding potential, the current drops until at a certain **stopping potential** V_s , all current ceases. Evidently V_s is given by[†]

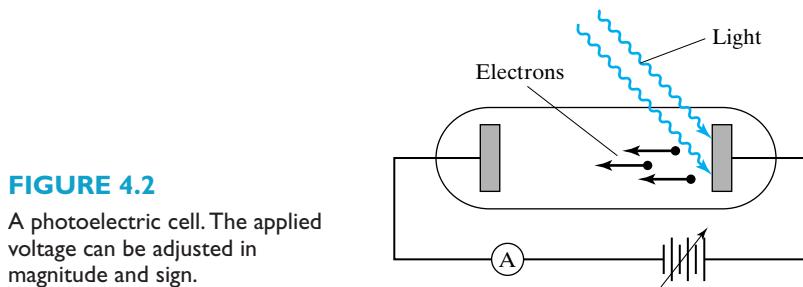
$$V_s e = K_{\max} \quad (4.3)$$

*Since some of our colleagues have expressed shock at our view that blackbody radiation is better skipped over at this level, we would like to mention two articles by distinguished physicists advocating the same view as ours: Donald Holcomb, *American Journal of Physics*, vol.61, p. 969 (1993) and Hans Bethe, *ibid*, p. 972.

[†]If the two electrodes are made of different metals, there is a small complication that arises because the two metals attract electrons differently. (This causes the so-called contact potential of the two metals.) To avoid this complication, we assume that the two electrodes are made of the same metal.



128 Chapter 4 • Quantization of Light



where K_{\max} denotes the maximum kinetic energy of the ejected electrons. Thus by measuring the stopping potential V_s one can find K_{\max} .

When the apparatus of Fig. 4.2 is used to investigate the numbers and kinetic energies of the electrons ejected, two important facts emerge:

1. If the intensity of the incident light is increased, the number of ejected electrons increases (as one might expect), but quite unexpectedly, *their kinetic energy does not change at all*.
2. If the frequency f of the incident light is reduced below a certain critical frequency f_0 *no electrons are ejected* however intense the light may be.

Neither of these results is consistent with the classical view of electromagnetic waves as a continuous distribution of oscillating electric and magnetic fields. According to this view, an increased intensity means increased field strengths, which should surely eject some electrons with increased kinetic energy; and (again in the classical view) there is no reasonable way to explain why low-frequency fields (however strong) should be unable to eject any electrons.

Einstein proposed that as a natural extension of Planck's ideas, one should assume that "the energy in a beam of light is not distributed continuously through space, but consists of a finite number of energy quanta, which are localized at points, which cannot be subdivided, and which are absorbed and emitted only as whole units." The energy of a single quantum (or photon as we would now say) he took to be hf . (These proposals go beyond — but include — Planck's assumption that light is emitted in multiples of hf .) Since it seemed unlikely that two photons would strike one electron, Einstein argued that each ejected electron must be the result of a single photon giving up its energy hf to the electron. With these assumptions, both of the properties mentioned above are easily explained as follows:

If the intensity of light is increased, then, according to Einstein's assumptions, the number of photons is increased, but the energy hf of an individual photon is unchanged. With more photons, more electrons will be ejected. But since each photon has the same energy, each ejected electron will be given the same energy. Therefore, K_{\max} will not change, and point (1) is explained.

Point (2) is equally easy: For any given metal, there is a definite minimum energy needed to remove an electron. This minimum energy is called the **work function** for the metal and is denoted ϕ . If the photon energy hf is less than ϕ , no photons will be able to eject any electrons; that is, if f is less than a critical frequency f_0 given by

$$hf_0 = \phi \quad (4.4)$$

no electrons will be ejected, as observed.

Einstein carried this reasoning further to make a quantitative prediction. If the frequency f is greater than f_0 , each ejected electron should have gained energy hf from a photon but lost ϕ or more in escaping from the metal. Thus, by conservation of energy, its kinetic energy on emerging should be $hf - \phi$, or less. Therefore,

$$K_{\max} = hf - \phi \quad (4.5)$$

That is, the observed maximum energy of the ejected electrons should be a linear function of the frequency of the light, and the slope of this function should be Planck's constant, h . The experimental test of this prediction proved difficult but was successfully carried out by Robert Millikan, one of whose trials (1916) is shown in Fig. 4.3. It will be seen that Millikan's data are a beautiful fit to the expected straight line. In particular, by measuring the slope, Millikan was able to determine h and obtained a value in agreement with that found by Planck.

If Planck and Einstein were right that light is quantized (as they were), the question naturally arises as to why this quantization had not been observed sooner. The answer is that, by everyday standards, the energy hf of a single photon is very small. Thus the number of photons in a normal beam of light is enormous, and the restriction of the energy to integer multiples of hf is correspondingly unimportant, as the following example illustrates.

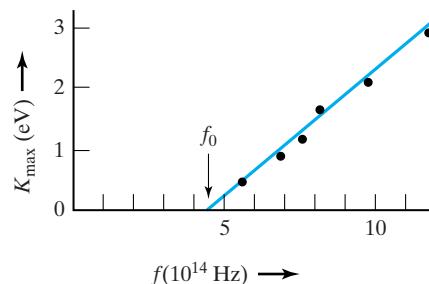


FIGURE 4.3
Millikan's data for K_{\max} as a function of frequency f for the photoelectric effect in sodium.

Example 4.1

What is the energy of a typical visible photon? About how many photons enter the eye per second when one looks at a weak source of light such as the moon, which produces light of intensity about 3×10^{-4} watts/m²?

The wavelength of visible light is between 400 and 700 nm. Thus we can take a typical visible wavelength to be

$$\lambda \approx 550 \text{ nm} \quad (4.6)$$

The energy of a single photon is

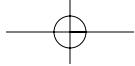
$$E = hf = \frac{hc}{\lambda} \quad (4.7)$$

Before we evaluate this, it is useful to note that the product hc enters into many calculations and is a useful combination to remember. Since h has the

Robert Millikan
(1868–1953, American)



As an undergraduate at Oberlin College, Millikan studied Greek, but he went on to study physics at Columbia and became that institution's first PhD in physics in 1895. After working with Planck in Germany, he became a professor at the University of Chicago and later at CalTech. Although he won the 1923 Nobel Prize in physics mainly for his beautiful measurement of the electron's charge (around 1910), the citation also mentioned his verification of Einstein's equation for the photoelectric effect.



130 Chapter 4 • Quantization of Light

dimension energy \times time, hc has the dimension energy \times length and is conveniently expressed in eV \cdot nm, as follows:

$$hc = (6.63 \times 10^{-34} \text{ J} \cdot \text{s}) \times (3.00 \times 10^8 \text{ m/s}) \times \frac{1 \text{ eV}}{1.60 \times 10^{-19} \text{ J}}$$

$$= 1.24 \times 10^{-6} \text{ eV} \cdot \text{m}$$

or

$$hc = 1240 \text{ eV} \cdot \text{nm} \quad (4.8)$$

Putting numbers into (4.7), we find the energy for a typical visible photon to be

$$E = \frac{hc}{\lambda} \approx \frac{1240 \text{ eV} \cdot \text{nm}}{550 \text{ nm}} \approx 2.3 \text{ eV} \quad (4.9)$$

On the atomic level this energy is significant, but by everyday standards it is extremely small.

When we look at the moon, the energy entering our eye per second is given by IA , where I is the intensity ($I \approx 3 \times 10^{-4} \text{ W/m}^2$) and A is the area of the pupil ($A \approx 3 \times 10^{-5} \text{ m}^2$ if we take the diameter of the pupil to be about 6 mm). Thus the number of photons entering our eye per second is

$$\begin{aligned} \text{number of photons per second} &= \frac{IA}{E} \\ &\approx \frac{(3 \times 10^{-4} \text{ W/m}^2) \times (3 \times 10^{-5} \text{ m}^2)}{(2.3 \times 1.6 \times 10^{-19} \text{ J})} \\ &\approx 2.5 \times 10^{10} \text{ photons per second} \end{aligned}$$

This is such a large number that the restriction to integer numbers of photons is quite unimportant even for this weak source.*

Example 4.2

The work function of silver is $\phi = 4.7 \text{ eV}$. What is the potential V_s needed to stop all electrons when ultraviolet light of wavelength $\lambda = 200 \text{ nm}$ shines on silver?

The energy of the UV photon is

$$hf = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{200 \text{ nm}} = 6.2 \text{ eV}$$

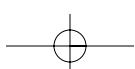
The stopping potential V_s is given by $V_s e = K_{\max}$, where K_{\max} is given by the Einstein equation (4.5). Thus

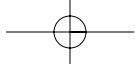
$$V_s e = K_{\max} = hf - \phi = (6.2 - 4.7) \text{ eV} = 1.5 \text{ eV}$$

or

$$V_s = 1.5 \text{ volts}$$

*Note, however, that the individual receptors in the eye are extraordinarily sensitive, being able to detect just a few photons per second.





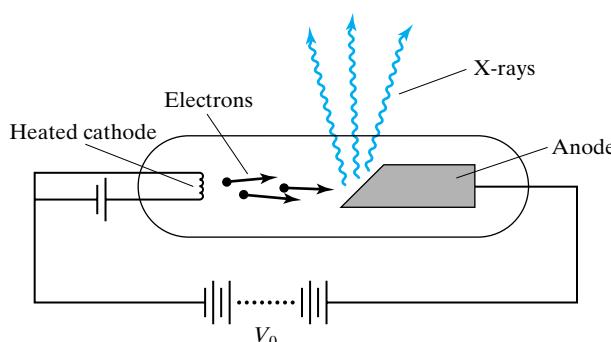
4.4 X-rays and Bragg Diffraction

The ideas that light was quantized and that the quantum of light should be regarded as a particle were slow to gain wide acceptance. In fact, the name “photon,” which recognizes these ideas, was not coined until 1926 (by the American chemist Gilbert Lewis). Probably the decisive event was the experiment (1923) in which Arthur Compton showed that photons carry momentum as well as energy and are subject to the same conservation laws of energy and momentum as other more familiar particles. Compton’s experiment used X-ray photons, and we therefore use the next two sections to describe X-rays before taking up the Compton effect itself.

X-rays are electromagnetic radiation whose wavelength is in or near the range from 1 nm down to 0.001 nm, at least 500 times shorter than visible wavelengths. Since shorter wavelengths mean higher photon energies, X-ray photons are much more energetic than visible photons, with energies of 1 keV or more. X-rays were discovered in 1895 by the German physicist Wilhelm Roentgen, who found that when energetic electrons were fired into a solid metal target, a very penetrating radiation was produced. Unable to identify this radiation, he gave it the name “X-rays.”

Figure 4.4 is a sketch of a modern medical X-ray tube (whose basic principles are actually quite similar to those of Roentgen’s original arrangement). Two electrodes are enclosed in an evacuated glass tube. Electrons are ejected from the heated cathode on the left. A potential difference of several thousand volts between the cathode and anode accelerates the electrons, which therefore acquire several keV of kinetic energy (and hence speeds of 0.1c and more). X-rays are produced when the electrons crash into the anode and are brought to an abrupt stop. It is found that most of the rays are emitted near 90° to the electrons’ path, and for this reason the anode is tilted to encourage the X-rays to exit in one direction, as shown.

The ability of X-rays to penetrate solids of low density was put to medical use within a few months of their discovery, but the problem of identifying what X-rays really were took longer. It was known that an accelerating electric charge produces electromagnetic waves. (For example, the oscillating charges in a radio antenna produce the long wavelength radiation that we call radio waves.) Thus it was perfectly reasonable to suppose that X-rays were electromagnetic waves produced by the enormous deceleration of the electrons stopping in the anode. (Radiation produced in this way is called **bremsstrahlung**, the German word for “braking radiation.”) The problem was to verify that X-rays really were waves, and the difficulty, as we now know, was that their wavelength is so very short.



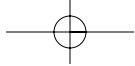
Arthur Compton (1892–1962, American)



After getting his PhD at Princeton, Compton studied with Rutherford at Cambridge. His observations (1923) of the scattering of electromagnetic radiation by atoms showed clearly that photons should be regarded as particles that carry energy and momentum. These studies earned him the 1927 Nobel Prize in physics. In the 1930s he worked on cosmic rays and helped establish that cosmic rays are charged particles, not (as advocated by Millikan) electromagnetic radiation.

FIGURE 4.4

An X-ray tube. The kinetic energy of the electrons hitting the anode is V_0e , which is, therefore, the maximum possible energy of the X-ray photons.



132 Chapter 4 • Quantization of Light

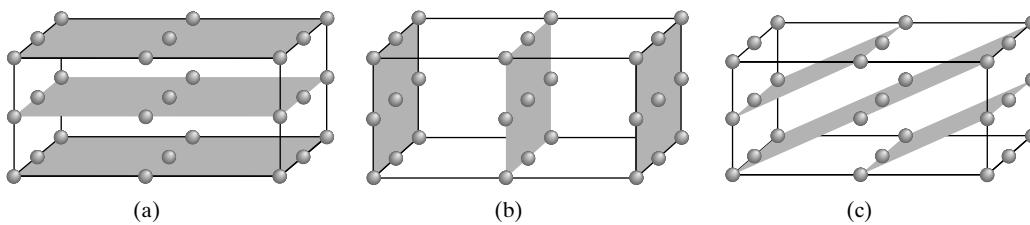


FIGURE 4.5

The atoms of any one crystal define many different sets of equally spaced parallel planes, each containing many atoms. Three such sets of planes are shown shaded.

One of the most effective ways to show that something is a wave and to measure its wavelength is to pass it through a diffraction grating and to observe the resulting interference pattern. In an efficient grating the slits must be spaced regularly with a separation of the same order as the wavelength. Thus a good grating for visible light has spacing of order 1000 nm, but a diffraction grating for X-rays would require spacing of order 0.1 nm — not something that could be easily made. In 1912 the German physicist Max von Laue suggested that since the atoms in a crystal are spaced regularly with separations of order 0.1 nm, it should be possible to use a crystal as a kind of three dimensional grating for X-rays. This proved correct, and Laue and his assistants quickly established that X-rays were waves, with wavelengths of order 0.1 nm.

The use of crystals as X-ray diffraction gratings was developed by the English physicists W. L. Bragg and his father, W. H. Bragg, and is often called **Bragg diffraction** (or Bragg scattering or Bragg reflection). The technique was historically important in the study of X-rays and is even more important today in the study of crystal structures. To understand the Braggs' analysis, we first note that we can think of a crystal as a large number of regularly spaced identical parallel planes, each containing many regularly spaced atoms as in Fig. 4.5(a). [These crystal planes can be chosen in several ways; two other possibilities are shown in Fig. 4.5(b) and (c).] Let us consider an electromagnetic wave approaching the planes of Fig. 4.5(a), with glancing angle θ , as shown in Fig. 4.6(a). (In Bragg diffraction, the incident direction is traditionally specified by the glancing angle measured up from the plane, rather than the angle down from the normal as in optics.) When the wave strikes the crystal, each

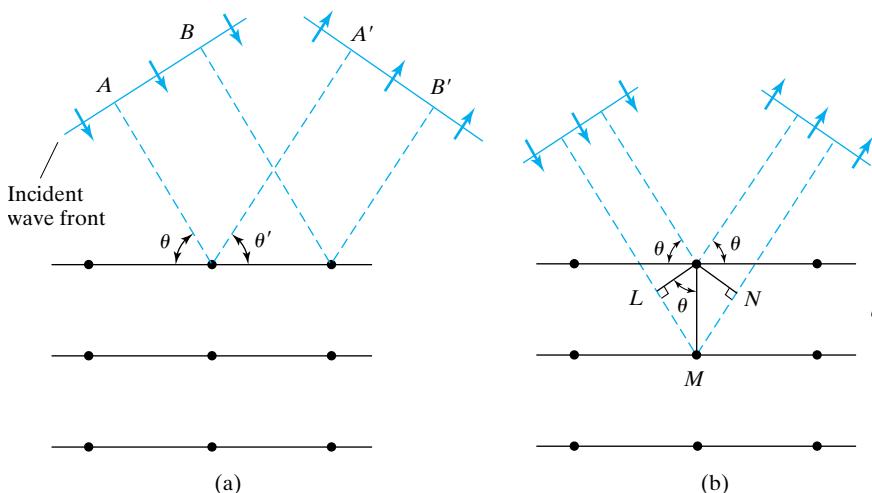


FIGURE 4.6

Side view of waves striking the crystal planes of Fig. 4.5(a).

- (a) Waves scattered by atoms in a single plane are in phase if the path lengths AA' and BB' are equal; this requires that $\theta = \theta'$. (b) The path difference for waves scattered from two successive planes is the distance $LMN = 2d \sin \theta$.

Section 4.4 • X-rays and Bragg Diffraction **I 33**

atom will scatter some of the radiation, and we will observe diffraction maxima in those directions where all of the scattered waves are in phase. If we consider first waves scattered by atoms in a single plane, all of the scattered waves will be in phase in the direction given by the familiar law of reflection [Fig. 4.6(a)]

$$\theta = \theta' \quad (4.10)$$

(This is why Bragg diffraction is often called Bragg reflection.)

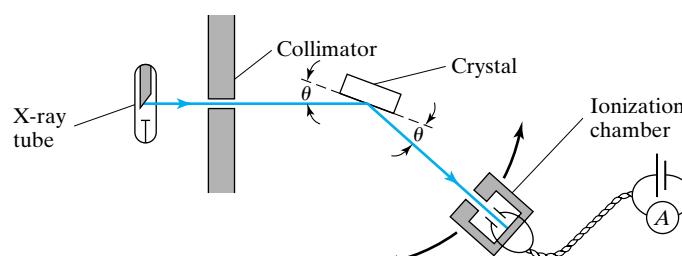
Let us next consider the waves scattered by atoms in two adjacent planes, a perpendicular distance d apart. It is easy to see, as in Fig. 4.6(b), that the path difference for these two waves is $2d \sin \theta$. Thus the waves from adjacent planes will be in phase provided that $2d \sin \theta$ is an integral multiple of the wavelength λ :

$$2d \sin \theta = n\lambda \quad (4.11)$$

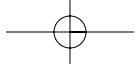
where n is any integer $1, 2, 3, \dots$ and is known as the order of the diffraction maximum. In many applications the maxima with $n > 1$ are relatively weak, so that only $n = 1$ is important.

The condition (4.11) is called the **Bragg law**. In any direction for which both (4.10) and (4.11) are satisfied, the waves from all atoms in the crystal will be in phase, and a strong maximum will be observed. This result can be used in several ways. For many simple crystals the spacing, d , of the planes can be calculated from knowledge of the density of the crystal and the mass of the atoms. If monochromatic X-rays (that is, X-rays of a single wavelength λ) are fired at a crystal of known spacing, the resulting pattern can be used, in conjunction with the Bragg law, to measure λ . If the X-rays contain a spread of wavelengths ("white" X-rays), the different wavelengths will give maxima in different directions, and one can use the crystal as an **X-ray spectrometer**, to find out what wavelengths are present and with what intensity. (A spectrometer is any device — like the familiar diffraction grating — that sorts and measures the different wavelengths in radiation.)

A simple X-ray spectrometer of the type used by the Braggs is shown in Fig. 4.7. The X-rays under study pass through a collimator (a small aperture to define their direction) and are reflected off a crystal whose plane spacing is known. The intensity I of the reflected rays is measured by a detector, such as an ionization chamber. (In this device, the X-rays pass through a gas between two plates at different potentials. The X-rays ionize the gas allowing a current to flow, whose magnitude is proportional to the X-ray intensity.) By rotating the crystal and detector, one can find the intensity I as a function of θ . By the Bragg law (4.11), this is equivalent to finding I as a function of λ , which is the required X-ray spectrum. By selecting just the X-rays scattered at one angle,

**FIGURE 4.7**

An X-ray spectrometer. X-rays are reflected off the crystal and detected by the ionization chamber. The crystal and chamber can both rotate in such a way that the two angles θ are always equal.



134 Chapter 4 • Quantization of Light

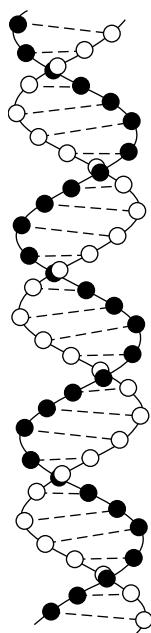


FIGURE 4.8

The double helix of the DNA molecule was inferred by Francis Crick and James Watson from X-ray studies. Each circle represents a base, comprising about 30 atoms, and the order of these bases carries genetic information. (The bases in the two strands are shaded differently only to emphasize the helical structure.)

one can obtain a monochromatic beam of X-rays for use in some other experiment, in which case we would call the arrangement an X-ray monochromator.

Once the wavelength of X-rays is known, one can use the Bragg law to investigate unknown crystal structures. This use of X-rays is called X-ray crystallography and is an important tool in solid-state physics and molecular biology. Figure 4.8 shows a short section of the biological molecule DNA, whose structure was found from X-ray studies.

The diffraction pattern implied by the Bragg law (4.11) is often fairly complicated. The main reason is that, as mentioned in connection with Fig. 4.5, there are many different sets of crystal planes, with different orientations and different spacings, d . For each such set of planes, the Bragg condition implies certain definite directions of maximum intensity. The resulting pattern in a typical X-ray diffraction experiment is shown in Fig. 4.9.

A second complication in Bragg diffraction is that many solids are not a single monolithic crystal, but are instead a jumbled array of many microcrystals. When X-rays pass into such a polycrystal, only those microcrystals that are oriented correctly for the Bragg condition will produce constructive interference. The locus of these constructive directions for all such microcrystals is a cone, and the resulting pattern is a series of rings as illustrated in Fig. 4.10. These rings can be thought of as the result of rotating a single-crystal pattern like Fig. 4.9 about its center.

4.5 X-ray Spectra

With an X-ray spectrometer such as the one described in Section 4.4, it was possible to analyze the distribution of wavelengths produced in an X-ray tube. This kind of distribution is generally recorded as a *spectrum*; that is, as a graph of intensity as a function of wavelength or frequency. Two typical spectra,

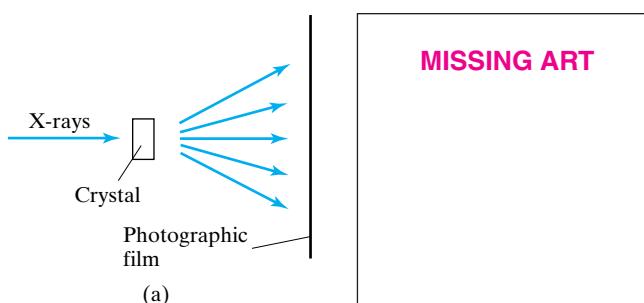


FIGURE 4.9

(a) One possible arrangement for observing X-ray diffraction with a single crystal, (b) The X-ray diffraction pattern produced by a crystal of the compound $\text{P}_2\text{S}_2(\text{NC}_6\text{H}_5)_2$ (NHC_6H_5)₂.

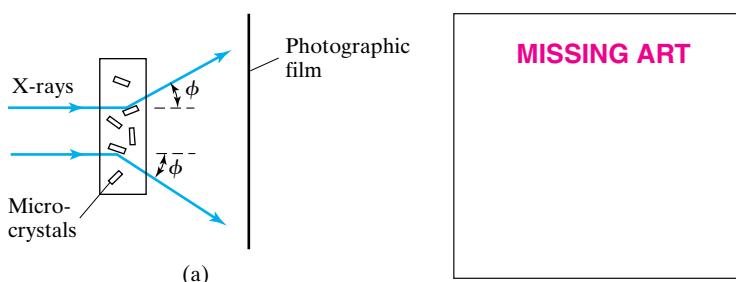
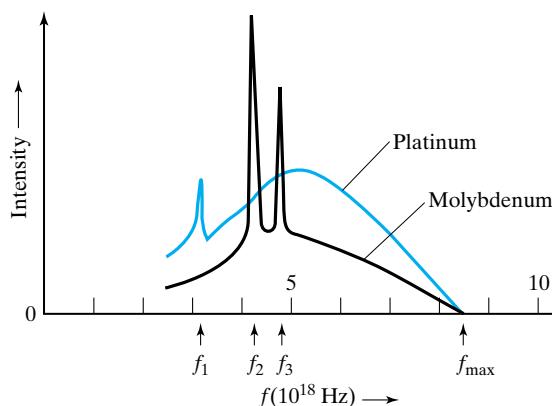


FIGURE 4.10

X-ray diffraction with a polycrystal. (a) Constructive interference occurs only for those microcrystals, that are oriented correctly for the Bragg condition; the locus of all constructive directions is a cone whose angle is shown as ϕ . (b) The resulting pattern for X-rays with $\lambda = 0.07 \text{ nm}$ fired through an aluminum foil.

**FIGURE 4.11**

X-ray spectra produced by platinum and molybdenum anodes, both made with an accelerating potential of 35 kV. Note how both spectra terminate abruptly at the same frequency, f_{\max} .

made with the same accelerating potential but with different anode metals, are sketched in Fig. 4.11, where the intensity is shown as a function of frequency.

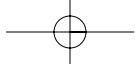
According to the classical theory of bremsstrahlung (the braking radiation produced by a decelerating charge), the X-rays should be produced in a broad spread of frequencies, such that the intensity varies smoothly with f and drops slowly to zero at high frequencies. The observed spectra in Fig. 4.11 contradict this prediction in two important ways. First, each curve shows one or more tall sharp spikes superposed on an otherwise smooth background. These spikes indicate that an appreciable fraction of the radiation produced in either metal is produced at certain isolated frequencies (marked f_1 for platinum and f_2, f_3 for molybdenum in the figure). Similar sharp spikes appear with whatever metal is used for the anode, but they occur at different frequencies that are characteristic of the metal concerned. For this reason, the X-rays at these frequencies are called **characteristic X-rays**.

Classical physics has no explanation for the characteristic X-rays, which are produced by a mechanism quite different from the bremsstrahlung of the smooth background curves in Fig. 4.11. We will return to the characteristic X-rays in Section 5.9, where we will see that they are the result of the quantization of atomic energy levels.

The second important feature of the spectra shown in Fig. 4.11 is that both drop abruptly to zero at a certain maximum frequency f_{\max} , which is the same for both metals. This phenomenon, which lacks any classical explanation, is easily explained if we recognize, first, that X-rays are quantized with energy hf and, second, that the energy of each quantum is supplied by one of the electrons striking the anode in the X-ray tube. These electrons have kinetic energy $K = V_0e$, where V_0 is the accelerating voltage of the tube. Since the most energy an electron can possibly give up is K , no X-ray photons can be produced with energy hf greater than K . Thus the maximum frequency produced must satisfy

$$hf_{\max} = K = V_0e \quad (4.12)$$

This result shows that f_{\max} varies in proportion to the accelerating voltage V_0 , but is the same for all anode materials. It is called the **Duane-Hunt law** after its discoverers William Duane and F. L. Hunt, who used it to obtain a value for Planck's constant h . Their result was in excellent agreement with Planck's value, furnishing "strong evidence in favor of the fundamental principle of the quantum hypothesis."



136 Chapter 4 • Quantization of Light

Example 4.3

The spacing of one set of crystal planes in common salt (NaCl) is $d = 0.282 \text{ nm}$. A monochromatic beam of X-rays produces a Bragg maximum when its glancing angle with these planes is $\theta = 7^\circ$. Assuming that this is the first-order maximum ($n = 1$), find the wavelength of the X-rays. What is the minimum possible accelerating voltage, V_0 , that produced the X-rays?

From the Bragg law (4.11), with $n = 1$, we find that

$$\begin{aligned}\lambda &= 2d \sin \theta = 2 \times (0.282 \text{ nm}) \times \sin 7^\circ \\ &= 0.069 \text{ nm}.\end{aligned}$$

The Duane–Hunt law requires that the electrons' kinetic energy V_0e in the X-ray tube must be at least equal to the energy hf of the X-ray photons. Therefore,

$$V_0e \geq hf = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{0.069 \text{ nm}} = 18,000 \text{ eV}$$

or

$$V_0 \geq 18,000 \text{ volts}$$

4.6 The Compton Effect

When a beam of light is fired at a system of charges, like an atom or a single electron, some of the beam is scattered in various directions. The classical theory of such scattering is straightforward: The oscillating electric field of the incident light causes the charges to oscillate, and the oscillating charges then radiate secondary waves in various directions. The angular distribution of these scattered waves depends on the detailed arrangement of the target charges, but one prediction of the classical theory is common to all targets: The frequency f of the scattered waves must be the same as that of the oscillating charges, which in turn must be the same as the incident frequency f_0 . Thus, in the classical view, the scattered and incident frequencies are necessarily the same,

$$f = f_0 \quad (4.13)$$

Numerous experiments with visible light and preliminary observations with X-rays had all seemed to confirm this prediction.

Starting in 1912, however, there appeared various reports that when high frequency X-rays were scattered off electrons, the scattered frequency f was less than f_0 ,

$$f < f_0 \quad (4.14)$$

This claim was so surprising that it was not taken seriously at first. But in 1923 the American physicist Arthur Compton published two papers in which he argued that if light is quantized, one should *expect* to find that $f < f_0$; and he

Section 4.6 • The Compton Effect 137

reported experiments that showed that for X-rays scattering off electrons, f is less than f_0 .

Compton argued that if photons carry energy, they should also carry momentum. This momentum p should be related to the energy E by the “Pythagorean relation” (2.23)

$$E^2 = (pc)^2 + (mc^2)^2 \quad (4.15)$$

except that since photons travel with speed c , they must have $m = 0$, and hence satisfy

$$E = pc \quad (4.16)$$

as discussed in Section* 2.8. Since $E = hf$, this implies that

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda}. \quad (4.17)$$

He proposed further that when radiation is scattered by electrons, each scattered photon results from a collision with a single electron, and that the ordinary rules of conservation of energy and momentum hold in this collision.

Compton's assumptions immediately explain the observed shift in frequency of the scattered X-rays: When a photon (of frequency f_0) strikes a stationary free electron, the electron must recoil. Since the electron gains energy in this process, the photon must lose energy. Therefore, the photon's final energy hf is less than the original hf_0 , and $f < f_0$, as observed.

In fact, Compton used conservation of energy and momentum to predict the scattered frequency f as a function of the scattering angle θ of the X-rays. Before we describe this calculation, we should emphasize that Compton did not, of course, have a target of stationary free electrons. In fact, his target electrons were the electrons in the carbon atoms of a graphite block. Thus his electrons were moving (in their atomic orbits), and because of the binding forces in the atoms, they were not perfectly free to recoil when struck by the photons. Fortunately, neither of these complications is important. X-rays have wavelengths of order 0.1 nm (Compton's had $\lambda = 0.07$ nm) and energies of order

$$E = hf = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{0.1 \text{ nm}} \approx 10^4 \text{ eV}$$

The kinetic energies of the outer electrons in an atom are a few eV, and the energy needed to remove them is of the same order. These energies are negligible compared to the incident photons' energies, and we can, therefore, treat the target electrons as if they were at rest and free.

Let us now consider a photon of energy $E_0 = hf_0$ and momentum \mathbf{p}_0 approaching a stationary electron with rest energy mc^2 and momentum zero. After the encounter, we suppose that the photon has an energy $E = hf$ and a momentum \mathbf{p} that makes an angle θ with \mathbf{p}_0 , as in Fig. 4.12. The electron recoils

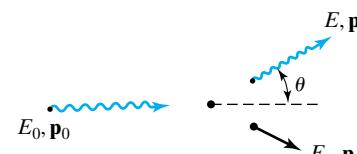
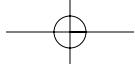


FIGURE 4.12

Compton scattering, in which a photon scatters off a free electron.

*If you did not read Chapters 1 and 2 on relativity, then we must ask you to accept this relation between the momentum and energy of electromagnetic radiation. In fact the relation (4.16) was known even before the advent of relativity, since it can be derived from classical electromagnetic theory. Either way, if you accept (4.16), you are ready to understand Compton's arguments.



138 Chapter 4 • Quantization of Light

with total energy E_e , and momentum \mathbf{p}_e . By conservation of energy and momentum, we expect that

$$E_e + E = mc^2 + E_0 \quad (4.18)$$

and

$$\mathbf{p}_e + \mathbf{p} = \mathbf{p}_0 \quad (4.19)$$

(We treat the electron relativistically so that our analysis will apply even at very high energies.) We can eliminate the electron's energy and momentum from these two equations and find the photon's final energy E in terms of the initial E_0 , and hence f in terms of f_0 .

We first solve (4.18) for E_e , to give

$$E_e = mc^2 + E_0 - E \quad (4.20)$$

The photon energies can be rewritten using (4.16) as $E_0 = p_0c$ and $E = pc$; and the electron's energy E_e is given by the Pythagorean relation (4.15). Canceling a common factor of c , we find that (4.20) implies that

$$\sqrt{p_e^2 + (mc)^2} = mc + p_0 - p \quad (4.21)$$

We can solve (4.19) to give $\mathbf{p}_e = \mathbf{p}_0 - \mathbf{p}$ and hence

$$\begin{aligned} p_e^2 &= \mathbf{p}_e \cdot \mathbf{p}_e = (\mathbf{p}_0 - \mathbf{p}) \cdot (\mathbf{p}_0 - \mathbf{p}) \\ &= p_0^2 + p^2 - 2\mathbf{p}_0 \cdot \mathbf{p} \\ &= p_0^2 + p^2 - 2p_0p \cos \theta. \end{aligned} \quad (4.22)$$

We can next substitute (4.22) for p_e^2 into (4.21). After squaring both sides and canceling several terms, we find that

$$mc(p_0 - p) = p_0p(1 - \cos \theta) \quad (4.23)$$

or

$$\frac{1}{p} - \frac{1}{p_0} = \frac{1}{mc}(1 - \cos \theta) \quad (4.24)$$

From (4.24) we can find the scattered photon's momentum p ; and from this we can calculate its energy $E = pc$, its frequency $f = E/h$, or its wavelength $\lambda = c/f$. The result is most simply expressed in terms of λ since according to (4.17), the photon momentum is $p = h/\lambda$ so that $1/p = \lambda/h$ (and similarly for p_0). Thus (4.24) implies that

$$\Delta\lambda \equiv \lambda - \lambda_0 = \frac{h}{mc}(1 - \cos \theta) \quad (4.25)$$

This result, first derived by Compton, gives the increase in a photon's wavelength when it is scattered through an angle θ . Since $\Delta\lambda \geq 0$, the wavelength is always *increased* and the frequency *decreased*, as anticipated. The shift in wavelength, $\Delta\lambda$, is zero at $\theta = 0$ and increases as a function of θ , to a maximum

of $2h/mc$ at $\theta = 180^\circ$ as sketched in Fig. 4.13. The magnitude of the shift is determined by the quantity*

$$\frac{h}{mc} = \frac{hc}{mc^2} = \frac{1240 \text{ eV} \cdot \text{nm}}{0.511 \text{ MeV}} = 0.00243 \text{ nm} \quad (4.26)$$

In Compton's experiment the incident wavelength was $\lambda_0 = 0.07 \text{ nm}$. Thus the predicted shift ranged up to 7% of the incident λ_0 , a shift that Compton's spectrometer could certainly detect.

Compton measured the scattered wavelength at four different angles θ and found excellent agreement with his prediction (4.25). This agreement gave "very convincing" support to the assumptions on which (4.25) was based, that photons carry momentum as well as energy and can be treated like particles, subject to the ordinary laws of conservation of energy and momentum.

Two final comments on Compton's formula and his experimental results: First, it is clear from (4.25) that the change of wavelength $\Delta\lambda$ is independent of the wavelength itself. Thus we should see the same change $\Delta\lambda$ for visible light as for X-rays. Why, then, is no shift observed with visible light? The answer is simple: The wavelength of visible light (400 to 700 nm) is about 5000 times larger than that of Compton's X-rays. Thus the fractional change $\Delta\lambda/\lambda_0$ is about 5000 times smaller for visible light than for X-rays and is in practice unobservable.[†] Quite generally, the longer the incident wavelength, the less important Compton's predicted shift will be.

Second, Compton found that at each angle *some* of the scattered X-rays had the wavelength given by his formula (4.25), but that some had λ equal to the incident wavelength λ_0 . This is clearly visible in Fig. 4.14, which shows Compton's spectrum of X-rays scattered at $\theta = 135^\circ$. The explanation of these unshifted X-rays is also simple, at least qualitatively. The derivation of (4.25) assumed a collision between the incident photon and a free electron (free, that is, to recoil when struck by the photon). As already discussed, this assumption is certainly good for the outer atomic electrons that are weakly bound. But the inner electrons are very tightly bound to the atom. (Several hundred eV is needed to remove an inner electron from a carbon atom.) Thus if a photon interacts with either an inner electron or the atomic nucleus, it may not detach an electron at all and the whole atom may recoil as a unit. In this case we should treat the process as a collision between the photon and the whole atom, and the same analysis would lead to the same formula (4.25), but with the electron mass m replaced by the mass of the whole atom. This implies a change $\Delta\lambda$, which is about 20,000 times smaller (in the case of carbon). For these collisions the change of wavelength will appear to be zero.

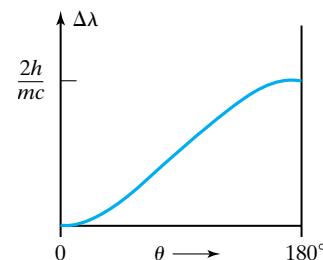


FIGURE 4.13

Increase, $\Delta\lambda$, in the wavelength of photons in Compton scattering. Note that $\Delta\lambda$ is zero at $\theta = 0$ and rises to a maximum of $2h/mc$ at $\theta = 180^\circ$.

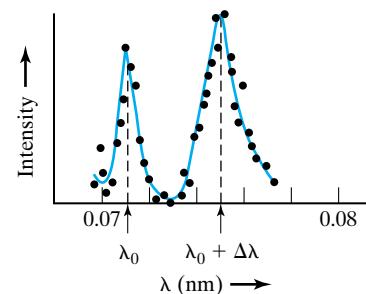


FIGURE 4.14

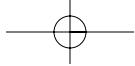
The spectrum of X-rays scattered at $\theta = 135^\circ$ off graphite, as measured by Compton; λ_0 is the incident wavelength (0.0711 nm) and $\Delta\lambda$ is the shift predicted by Compton's formula (4.25).

4.7 Particle-Wave Duality

Today all physicists accept that the photoelectric effect, the Compton effect, and numerous other experiments demonstrate beyond doubt the particle nature of light. But what about the many experiments that had established the

*This quantity is rather misleadingly called the *electron's Compton wavelength*. It is not, in any obvious sense, the wavelength of the electron; it is simply a quantity, with the dimensions of length, that determines the shift in wavelength of photons scattered off electrons.

[†]Even if this small shift were detectable, the Compton effect for visible light would be more complicated because the target electron cannot be considered as stationary and free if it is struck by a visible photon whose energy is only a few eV.



140 Chapter 4 • Quantization of Light

wave nature of light? Were these experiments somehow wrong? The answer is that both kinds of experiment are right, and that light exhibits wave properties *and* particle properties. In fact, both aspects of light are inextricably mixed in the two basic equations

$$E = hf \quad \text{and} \quad p = \frac{h}{\lambda} \quad (4.27)$$

since the energy E and momentum p refer to the particle nature of the photon, whereas the frequency f and wavelength λ are both wave properties.

We will see in Chapter 6 that ordinary particles like electrons and protons also show this **particle–wave duality**, and the first task of quantum theory is to reconcile these seemingly contradictory aspects of all the particles of modern physics. It turns out that the quantum theory of electrons, protons, and other massive particles is much simpler than the quantum theory of the massless photon. This is because massive particles can move at nonrelativistic speeds, whereas photons always move at speed c and are intrinsically relativistic. Thus for massive particles we can and will develop a nonrelativistic quantum mechanics, which is able to explain a wide range of phenomena in atomic, nuclear, and condensed matter physics. Relativistic quantum theory — in particular, the quantum theory of electromagnetic radiation, or *quantum electrodynamics* — is beyond the scope of this book. Fortunately, we can understand a large part of modern physics, armed with just the basic facts of quantum radiation theory, as summarized in the two equations (4.27).

CHECKLIST FOR CHAPTER 4

Concept	Details
Quantized	Can only occur in certain discrete amounts
Blackbody radiation	Radiation from a perfectly absorbing body — the phenomenon that caused Planck to suggest quantization of electromagnetic radiation
Planck's constant h	$h = 6.63 \times 10^{-34} \text{ J} \cdot \text{s}$
Photon energy	$E = hf$ (Secs. 4.2 and 4.3)
Photoelectric effect	Ejection of electrons from a metal by radiation $K_{\max} = hf - \phi$ (4.5)
Work function of a metal	ϕ = minimum energy to remove an electron
X-rays	Electromagnetic radiation with $0.001 \lesssim \lambda \lesssim 1 \text{ nm}$
Bragg diffraction	X-ray diffraction off crystals (Sec. 4.4)
Bragg law for X-ray diffraction	$2d \sin \theta = n\lambda$ (4.11)
X-ray spectra	Intensity of X-rays as a function of frequency (Sec. 4.5)
Duane–Hunt law for X-ray production	$hf_{\max} = V_0 e$ (4.12)
Compton effect	Scattering of photons by electrons (Sec. 4.6)
Compton formula	$\Delta\lambda = \frac{h}{mc}(1 - \cos \theta)$ (4.25)
Photon momentum	$p = h/\lambda$ (4.17)
Particle–wave duality	Photons (and all other particles) have both wavelike and particle-like properties

PROBLEMS FOR CHAPTER 4

SECTION 4.2 (Planck and Blackbody Radiation)

- 4.1** • The **intensity distribution function** $I(\lambda, T)$ for a radiating body at absolute temperature T is defined so that the intensity of radiation between wavelengths λ and $\lambda + d\lambda$ is

$$(\text{intensity between } \lambda \text{ and } \lambda + d\lambda) = I(\lambda, T) d\lambda$$

This is the power radiated per unit area of the body with wavelengths between λ and $\lambda + d\lambda$. The Planck distribution function for blackbody radiation is

$$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda k_B T} - 1} \quad (4.28)$$

where h is Planck's constant, c is the speed of light, and k_B is Boltzmann's constant. Sketch the behavior of this function for a fixed temperature for $0 < \lambda < \infty$. Explain clearly how you figured the trends of your graph.

[HINT: You should probably think about the two factors separately.]

- 4.2** • The Rayleigh–Jeans distribution function (which was based on classical ideas and turned out to be incorrect) is $I(\lambda, T) = 2\pi c k_B T / \lambda^4$. Show that for long wavelengths, this approximates the Planck formula (4.28). (Remember the Taylor series for e^x .) Sketch the two distribution functions. Can you explain why one might expect the classical result to be better at long wavelengths?

- 4.3** • It is often easier to characterize radiation by its frequency, rather than its wavelength. The Planck formula (4.28) is then written in terms of a function $I(f, T)$ defined so that $I(f, T) df$ is the intensity in the frequency interval from f to $f + df$. **(a)** Show that $I(f, T) = I(\lambda, T) |d\lambda/df|$. (The absolute value signs are needed to keep both distribution functions positive.) **(b)** Write down the Planck distribution function in terms of frequency f , and sketch its behavior at a fixed temperature T for $0 < \lambda < \infty$.

- 4.4** • The total intensity $I(T)$ radiated from a blackbody (at all wavelengths λ) is equal to the integral over all wavelengths, $0 < \lambda < \infty$, of the Planck distribution (4.28) (Problem 4.1). **(a)** By changing variables to $x = hc/\lambda k_B T$, show that $I(T)$ has the form

$$I(T) = \sigma T^4$$

where σ is a constant independent of temperature. This result is called **Stefan's fourth-power law**, after the Austrian physicist Josef Stefan. **(b)** Given that $\int x^3 dx / (e^x - 1) = \pi^4/15$, show that the Stefan–Boltzmann constant σ is $\sigma = 2\pi^5 k_B^4 / 15 h^3 c^2$. **(c)** Evaluate σ numerically, and find the total power radiated from a red-hot ($T = 1000$ K) steel ball of radius 1 cm. (Such a ball is well approximated as a blackbody.)

SECTION 4.3 (The Photoelectric Effect)

- 4.5** • Given that visible light has $400 < \lambda < 700$ nm, what is the range of energies of visible photons (in eV)? Ultraviolet (UV) radiation has wavelengths *shorter* than visible, while infrared (IR) wavelengths are *longer* than visible. What can you say about the energies of UV and IR photons?

- 4.6** • **(a)** A typical AM radio frequency is 1000 kHz. What is the energy in eV of photons of this frequency? **(b)** What is the energy of the photons in an FM signal of frequency 100 MHz?

- 4.7** • **(a)** X-rays are electromagnetic radiation with wavelengths much shorter than visible — of order 1 nm or less. What are the energies of X-ray photons? **(b)** Electromagnetic waves with wavelengths even shorter than X-rays are called γ rays (gamma rays) and are produced in many nuclear processes. A typical γ -ray wavelength is 10^{-4} nm (or 100 fm); what is the corresponding photon energy? Give your answers in keV or MeV, as appropriate.

- 4.8** • **(a)** Microwaves (as used in microwave ovens, telephone transmission, etc.) are electromagnetic waves with wavelength of order 1 cm. What is the energy of a typical microwave photon in eV? **(b)** The so-called 3-K radiation from outer space consists of photons of energy $k_B T$, where $T = 3$ K. What is the wavelength of this radiation?

- 4.9** • A typical chemical bond in a biological molecule has a strength of a few eV — let's say 4 eV to be specific. **(a)** Can low-intensity microwave radiation with wavelength 1 cm break a bond, thus causing a mutation in a DNA molecule? (Very high-intensity microwave radiation, such as is found inside a microwave oven, will *heat* biological tissue, thus causing burn damage.) **(b)** What is the minimum wavelength necessary for a photon to be able to break a 4-eV chemical bond? **(c)** What type of photon has this minimum wavelength? (Blue? IR? etc.)

- 4.10** • **(a)** Find the value of Planck's constant h in eV · sec. **(b)** This is a useful number to know when relating a photon's frequency to its energy. Use it to find the frequency of a 3-eV photon.

- 4.11** • The minimum frequency of radiation that can eject photoelectrons from a certain metal is 6×10^{14} Hz. What is the work function of this metal? (The result of Problem 4.10 will save a little trouble here.) What type of photon has this frequency? (Blue? Green? IR? etc.)

- 4.12** • The work function of cesium is 1.9 eV. This is the lowest value work function of any metal, and consequently, cesium is used in photomultiplier tubes, a kind of visible-light detector so sensitive that single photons can be readily detected. **(a)** What is the maximum wavelength of light that can eject photoelectrons from cesium? **(b)** If light with $\lambda = 500$ nm strikes cesium, what is the maximum kinetic energy of the ejected electrons?

142 Chapter 4 • Quantization of Light

4.13 • The longest wavelength of light that can eject electrons from potassium is $\lambda_0 = 560$ nm. **(a)** What is the work function of potassium? **(b)** If UV radiation with $\lambda = 300$ nm shines on potassium, what is the stopping potential V_s (the potential that just stops all the ejected electrons)?

4.14 •• A lightbulb that is rated at 60 W actually produces only about 3 W of visible light, most of the rest of the energy being infrared (or heat). **(a)** About how many visible photons does such a lightbulb produce each second? Use the average value $\lambda \approx 550$ nm. **(b)** If a person looks at such a bulb from about 10 ft away, about how many visible photons enter the eye per second? (When looking at a bright light, the pupil has a diameter of about 1 mm.) **(c)** By how many powers of 10 does this exceed the minimum detectable intensity, which is about 100 photons entering the eye per second?

4.15 •• The *solar constant* is the flux (energy per time per area) in radiation from the sun at the distance of the earth ($R = 150$ million km) and has the approximate value 1350 W/m^2 . **(a)** What is the sun's total energy output in Watts? **(b)** About half of the sun's output is in the visible range, with average wavelength about 550 nm. How many visible photons are coming out of the sun per second? **(c)** Estimate the number of visible photons hitting your face each second if you stand facing the sun directly.

4.16 •• The sun gets its energy from a sequence of nuclear reactions called the proton-proton cycle. These reactions take place near the sun's center and have the net effect



the particles involved being (in order) protons, helium nuclei, positrons, neutrinos, and high-energy photons (called γ rays). About 95% of the energy released in this composite reaction goes to the γ -ray photons, whose average wavelength is about 250 fm ($1 \text{ fm} = 10^{-15} \text{ m}$). **(a)** What is the total energy released in the above reaction? **(b)** Over the course of a few million years, these γ rays diffuse toward the surface and are slowly converted into visible photons. About how many visible photons are generated from one of the original γ 's?

4.17 •• Use Millikan's data from the photoelectric effect in sodium (Fig. 4.3) to get a rough value of Planck's constant h in $\text{eV} \cdot \text{s}$ and in $\text{J} \cdot \text{s}$.

4.18 •• The work function for tungsten is $\phi = 4.6 \text{ eV}$. **(a)** If light is incident on tungsten, find the critical frequency f_0 , below which no electrons will be ejected, and the corresponding wavelength λ_0 . Use the equation $K_{\max} = hf - \phi$ to find the maximum kinetic energy of ejected electrons if tungsten is irradiated with light with **(b)** $\lambda = 200 \text{ nm}$ and **(c)** 300 nm . Explain your answer to part (c).

SECTION 4.4 (X-Rays and Bragg Diffraction)

4.19 • Potassium chloride (KCl) has a set of crystal planes separated by a distance $d = 0.31 \text{ nm}$. At what glancing angle θ to these planes would the first-order Bragg maximum occur for X-rays of wavelength 0.05 nm ?

4.20 • When X-rays of wavelength $\lambda = 0.20 \text{ nm}$ are reflected off the face of a crystal, a Bragg maximum is observed at a glancing angle of $\theta = 17.5^\circ$, with sufficient intensity that it is judged to be first order. **(a)** What is the spacing d of the planes that are parallel to the face in question? **(b)** What are the glancing angles of all higher order maxima?

4.21 •• A student is told to analyze a crystal using Bragg diffraction. She finds that the ancient equipment has seized up and cannot turn to glancing angles below $\theta = 30^\circ$. She bravely persists and, using X-rays with $\lambda = 0.0438 \text{ nm}$, finds three weak maxima at $\theta = 36.7^\circ$, 52.8° , and 84.5° . What are the orders of these maxima, and what is the spacing of the crystal planes?

SECTION 4.5 (X-Ray Spectra)

4.22 • What is the shortest wavelength of X-rays that can be produced by an X-ray tube whose accelerating voltage is 10 kV ?

4.23 • What is the voltage of an X-ray tube that produces X-rays with wavelengths down to 0.01 nm but no shorter?

4.24 •• A monochromatic beam of X-rays produces a first order Bragg maximum when reflected off the face of an NaCl crystal with glancing angle $\theta = 20^\circ$. The spacing of the relevant planes is $d = 0.28 \text{ nm}$. What is the minimum possible voltage of the tube that produced the X-rays?

SECTION 4.6 (The Compton Effect)

4.25 • Use Compton's formula (4.25) to calculate the predicted shift $\Delta\lambda$ of wavelength at $\theta = 135^\circ$ for Compton's data shown in Fig. 4.14. What percent shift in wavelength was this?

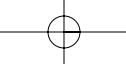
4.26 • Find the change in wavelength for photons scattered through 180° by free protons. Compare with the corresponding shift for electrons.

4.27 • A photon of 1 MeV collides with a free electron and scatters through 90° . What are the energy of the scattered photon and the kinetic energy of the recoiling electron?

4.28 •• Consider a head on, elastic collision between a massless photon (momentum \mathbf{p}_0 and energy E_0) and a stationary free electron. **(a)** Assuming that the photon bounces directly back with momentum \mathbf{p} (in the direction of $-\mathbf{p}_0$) and energy E , use conservation of energy and momentum to find p . **(b)** Verify that your answer agrees with that given by Compton's formula (4.25) with $\theta = \pi$.

- 4.29** •• Compton showed that an individual photon carries momentum, $p = E/c$, as well as energy, E . This momentum manifests itself in the radiation pressure felt by bodies exposed to bright light, as the following problem illustrates: A 100-W beam of light shines for 1000 s on a 1-g black object initially at rest in a frictionless environment. The object absorbs all the light of the beam. **(a)** Calculate the total energy and momentum of the photons absorbed by the blackbody. **(b)** Use conservation of momentum to find the body's final velocity. **(c)** Calculate the body's final kinetic energy; explain how this can be less than the original energy of the photons.
- 4.30** •• A student wants to measure the shift in wavelength predicted by Compton's formula (4.25). She finds that she cannot measure a shift of less than 5% and that she cannot conveniently measure at angles greater than $\theta = 150^\circ$. What is the longest X-ray wavelength that she can use and still observe the shift?
- 4.31** •• If the maximum kinetic energy given to the electrons in a Compton scattering experiment is 10 keV, what is the wavelength of the incident X-rays?
- 4.32** •• On a single graph, make plots of the Planck distribution function (4.28) for temperatures $T = 1000$ and 1500 K. Comment. (Where do the two graphs peak? How do the total radiated powers compare?)
- 4.33** ••• For a given temperature T , find the wavelength at which blackbody radiation is most intense; that is, find the value of λ for which the Planck function (4.28) is maximum. Show that $\lambda_{\max} \propto 1/T$, a result called **Wien's displacement law** ("displacement" because it specifies how λ_{\max} moves with temperature). (This is surprisingly hard. It may help to rewrite $I(\lambda, T)$ in terms of the variable $x = \lambda k_B T / hc$, and then find the value of x where I is maximum. This will lead you to a transcendental equation for x , which you can solve only with a computer. Some software can solve an equation of the form $f(x) = 0$ automatically, but you can also solve it graphically (to any desired accuracy) by making a plot of $f(x)$ and zooming in on the neighborhood of the zero.)

 **COMPUTER PROBLEMS**



Chapter 5

Quantization of Atomic Energy Levels

- 5.1 Introduction
- 5.2 Atomic Spectra
- 5.3 The Balmer–Rydberg Formula
- 5.4 The Problem of Atomic Stability
- 5.5 Bohr’s Explanation of Atomic Spectra
- 5.6 The Bohr Model of the Hydrogen Atom
- 5.7 Properties of the Bohr Atom
- 5.8 Hydrogen Like Ions
- 5.9 X-ray Spectra
- 5.10 Other Evidence for Atomic Energy Levels*

Problems for Chapter 5

*This section can be omitted without serious loss of continuity

5.1 Introduction

In Chapter 4 we described the discovery of the quantization of light, which made clear that classical electromagnetic theory is incorrect on the microscopic level. In the present chapter we describe the corresponding failure of classical mechanics when applied to microscopic systems. With hindsight, we can see that the evidence for this breakdown of classical mechanics goes back to the discovery of atomic spectra in the middle of the nineteenth century, as we describe in Sections 5.2 and 5.3. However, it was not until 1913 that any satisfactory explanation of atomic spectra was found — by the Danish physicist Niels Bohr — and it became clear that a substantial revision of classical mechanics was required.

Bohr’s work was originally prompted by a problem concerning the stability of Rutherford’s model of the atom (a problem we describe briefly in Section 5.4), but he quickly found that many properties of atomic spectra could easily be explained if one assumed that the total energy of the electrons in an atom is quantized. This quantization of atomic energies has no classical explanation, and to account for it, Bohr developed a mechanics that is now called the **Bohr model**, or the old quantum theory. As Bohr was well aware, his ideas were not really a complete theory, and they have now been superseded by modern quantum mechanics. Nevertheless, Bohr’s ideas were a crucial step in the development of modern quantum mechanics and were correct in several important respects. For these reasons, we describe the Bohr theory and some of its successes in the last six sections of this chapter.

5.2 Atomic Spectra

Perhaps the most famous spectrum of all time was the one discovered in 1666 by Isaac Newton, who shone a narrow beam of white light through a glass prism, producing the well-known ribbon of rainbow colors, as shown in the first spectrum at the front of this book. This established that what we perceive as white light is a mixture of different colors, or different wavelengths as we would now say.

In 1814 the German physicist Joseph von Fraunhofer discovered that when viewed more closely, the spectrum of sunlight is crossed by dark lines, like those in the second spectrum at the front of this book (though narrower and much more numerous). This showed that certain colors, or wavelengths, are missing from the light that reaches us from the sun. Today we know that this is because the gases in the sun's outer atmosphere absorb light at certain discrete wavelengths. The light with these wavelengths is therefore removed from the white light coming from deeper down, and this causes the dark lines observed by Fraunhofer.

By the middle of the nineteenth century it was known that all gases absorb light at wavelengths that are characteristic of the atoms and molecules they contain. For example, if white light is shone through a gas containing just one kind of atom, the gas will absorb certain wavelengths characteristic of that atom; if the transmitted light is then passed through a prism or diffraction grating, it will produce an **absorption spectrum** consisting of a bright ribbon of rainbow colors crossed with dark **absorption lines**, like Fraunhofer's. Furthermore, it is found that if the same gas is heated sufficiently, it will *emit* light. Moreover, the wavelengths of this emitted light are the same as those that the gas absorbed when illuminated with white light. If this emitted light is passed through a prism, it will produce an **emission spectrum**, consisting of bright **emission lines** against a dark background.

The absorption and emission spectra produced by atomic hydrogen are shown in color at the front of this book. (At ordinary temperatures hydrogen gas consists of H₂ *molecules*; to produce the atomic spectra shown, one must use gas that is heated enough — by an electric discharge, for example — to dissociate the molecules into atoms.) The same spectra are shown schematically in black and white in Fig. 5.1, where the pictures (a) and (c) show the emission and absorption spectra themselves, while (b) and (d) are the corresponding graphs of intensity against wavelength.

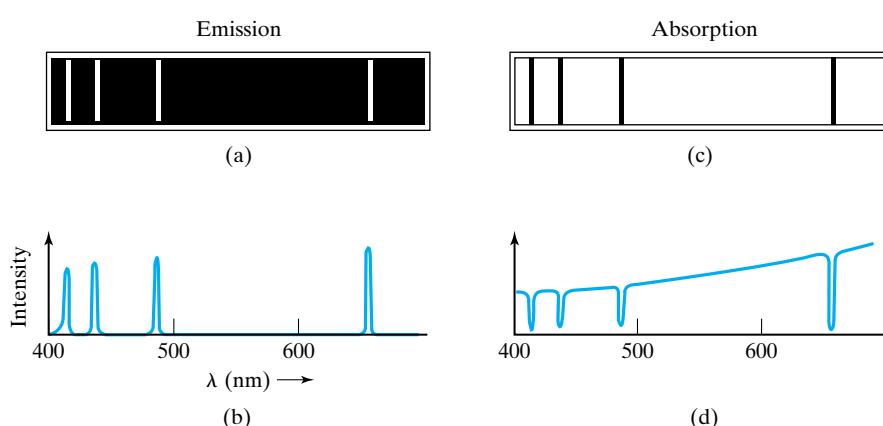
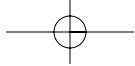


FIGURE 5.1

Emission and absorption spectra of atomic hydrogen. (a) The emission spectrum; the white stripes represent bright lines against a dark background. (b) The corresponding graph of intensity against wavelength, on which the spikes correspond to the bright lines of the spectrum itself (The relative intensities of the four lines depend on the temperature.) (c) The absorption spectrum, with dark lines against a bright background. (d) The corresponding graph, on which the dips correspond to the dark lines of the spectrum.



146 Chapter 5 • Quantization of Atomic Energy Levels

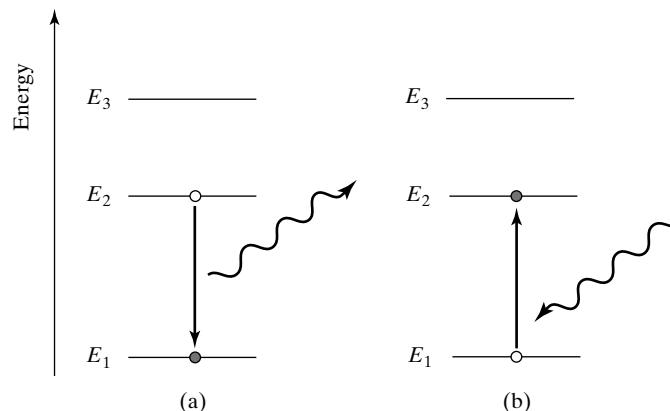
The atoms and molecules of any one chemical species emit and absorb light at wavelengths characteristic of that species. Thus emission and absorption spectra act like fingerprints, uniquely identifying the atom or molecule that produced them. By about 1870 spectroscopy had become a powerful tool of chemical analysis and had led to the discovery of several previously unknown elements. In particular, it was, and still is, the only way to determine the chemical composition of the sun, other stars, and interstellar matter.

Despite the many successful applications of spectroscopy in the nineteenth century, there was no satisfactory theory of atomic spectra. Classically, emission and absorption were easy to understand: An atom would be expected to emit light if some source of energy, such as collisions with other atoms, caused its electrons to vibrate and produce the oscillating electric fields that constitute light. Conversely, if light were incident on an atom, the oscillating electric field would cause the electrons to start vibrating and hence to absorb energy from the incident light. The observation that light is emitted and absorbed only at certain characteristic frequencies was understood to imply that the atomic electrons could vibrate only at these same frequencies; but no completely satisfactory classical model was ever found that could explain (let alone predict) these characteristic frequencies of vibration.

As we will see, Bohr's ideas (and likewise modern quantum mechanics) explain the characteristic spectra in a quite different way. First, the characteristic frequencies, f_α, f_β, \dots of light emitted by an atom imply that atoms emit photons with characteristic energies, $hf_\alpha, hf_\beta, \dots$ (This connection between frequency and energy was, of course, completely unknown in the nineteenth century.) These characteristic energies are explained by establishing that the total energy of the electrons in an atom is *quantized*, with discrete allowed values E_1, E_2, E_3, \dots , as illustrated in the **energy-level diagrams** of Fig. 5.2. An atom emits or absorbs light by making an abrupt jump from one energy state to another — for example, by changing from energy E_2 to E_1 , or vice versa. If $E_2 > E_1$ and the atom changes from E_2 to E_1 it must release the excess energy $E_2 - E_1$ and it does so in the form of a photon of energy $hf = E_2 - E_1$, as shown symbolically in Fig. 5.2(a); similarly, it can only change from E_1 to E_2 if it is supplied with energy $E_2 - E_1$ and one way this can happen is by absorption of a photon of energy $hf = E_2 - E_1$, as in Fig. 5.2(b). The characteristic energies of the photons emitted and absorbed by an atom are thus explained as the differences in the characteristic quantized energies of the atom. Before we explore Bohr's explanation of atomic spectra further, we relate a little more history.

FIGURE 5.2

The possible energies of an atom are found to be quantized, with discrete values E_1, E_2, E_3, \dots . In these energy-level diagrams, energy is plotted upward, and the allowed energies are represented as the rungs of a ladder. (a) If the atom is initially in level E_2 , it can drop to level E_1 by emitting a photon of energy $hf = E_2 - E_1$. (b) If it is initially in the level E_1 , it can absorb a photon of energy $hf = E_2 - E_1$, which will lift it to the level E_2 .



5.3 The Balmer–Rydberg Formula

The simplest of all atoms is hydrogen and it is therefore not surprising that the spectrum of atomic hydrogen was the first to be thoroughly analyzed. By 1885 the four visible lines shown in Fig. 5.1 had been measured very accurately by the Swedish astronomer and physicist Anders Ångström. These measurements were examined by a Swiss school teacher Johann Balmer, who found (in 1885) that the observed wavelengths fitted the formula

$$\frac{1}{\lambda} = R \left(\frac{1}{4} - \frac{1}{n^2} \right) \quad (5.1)$$

where R was a constant (with the dimension length⁻¹), which Balmer determined as

$$R = 0.0110 \text{ nm}^{-1} \quad (5.2)$$

and n was an integer equal to 3, 4, 5, and 6 for the four lines in question. Ångström had measured these wavelengths to four significant figures, and Balmer's formula fitted them to the same accuracy. Balmer guessed (correctly, as we now know) that such an excellent fit could not be a coincidence and that there were probably other lines given by other values of the integer n in the formula (5.1). For example, if we take $n = 7$, then (5.1) gives

$$\lambda = 397 \text{ nm}$$

a wavelength near the violet edge of the visible spectrum; with $n = 8, 9, \dots$, Eq. (5.1) predicts shorter wavelengths in the ultraviolet. One can imagine Balmer's delight when he discovered that several more lines had already been observed in the spectrum of hydrogen and that they were indeed given by his formula with $n = 7, 8, 9, \dots$.

We can rewrite Balmer's formula (5.1) in the form

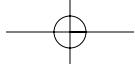
$$\frac{1}{\lambda} = R \left(\frac{1}{2^2} - \frac{1}{n^2} \right) \quad (n = 3, 4, 5, \dots) \quad (5.3)$$

It is tempting to guess that this is just a special case of the more general formula

$$\frac{1}{\lambda} = R \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (n > n', \text{ both integers}) \quad (5.4)$$

and that the spectrum of atomic hydrogen should contain all wavelengths given by all integer values of n' and n . Balmer himself had guessed that some such generalization might be possible, but the form (5.4) was apparently first written down by the Swedish physicist Johannes Rydberg, for whom (5.4) is usually called the **Rydberg formula** and R , the **Rydberg constant**. If we take $n' = 1$ and $n = 2$, for example, Rydberg's formula predicts

$$\lambda = 121 \text{ nm}$$



148 Chapter 5 • Quantization of Atomic Energy Levels

a wavelength well into the ultraviolet; with $n' = 3, n = 4$, we get

$$\lambda = 1870 \text{ nm}$$

in the infrared. In fact, all of the additional wavelengths predicted by (5.4) (with n' any integer other than Balmer's original value of 2) are either in the ultraviolet or the infrared. It was several years before any of these additional lines were observed, but in 1908 the German physicist Louis Paschen found some of the infrared lines with $n' = 3$, and in 1914 the American Theodore Lyman found some of the ultraviolet lines with $n' = 1$. Today, it is well established that the Rydberg formula (5.4) accurately describes all of the wavelengths in the spectrum of atomic hydrogen.

Example 5.1

Conventional spectrometers with glass components do not transmit ultraviolet light ($\lambda \leq 380 \text{ nm}$). Explain why none of the lines predicted by (5.4) with $n' = 1$ could be observed with a conventional spectrometer.

For the case $n' = 1, n = 2$, Eq. (5.4) predicts that

$$\frac{1}{\lambda} = R \left(\frac{1}{1} - \frac{1}{4} \right) = \frac{3}{4} R$$

and hence

$$\lambda = \frac{4}{3R} = \frac{4}{3 \times (0.0110 \text{ nm}^{-1})} = 121 \text{ nm}$$

as stated earlier. Similarly, for $n' = 1$ and $n = 3$, one finds that $\lambda = 102 \text{ nm}$, and inspection of (5.4) shows that the larger we take n , the *smaller* the corresponding wavelength, with $\lambda \rightarrow 91 \text{ nm}$ as $n \rightarrow \infty$. Therefore, all lines with $n' = 1$ lie in the range $91 < \lambda \leq 121 \text{ nm}$, well into the ultraviolet, and are unobservable with a conventional spectrometer.

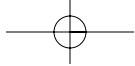
It is often convenient to rewrite the Rydberg formula (5.4) in terms of photon energies rather than wavelengths. Since the energy of a photon is* $E_\gamma = hc/\lambda$, we have only to multiply (5.4) by hc to give

$$E_\gamma = hcR \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (n > n', \text{ both integers}) \quad (5.5)$$

These are the energies of the photons emitted or absorbed by a hydrogen atom.

It is important to recognize that neither Balmer, Rydberg, nor anyone else prior to 1913 could explain *why* the formula (5.5) gave the spectrum of hydrogen. It was perhaps Bohr's greatest triumph that his theory *predicted* the Rydberg formula (5.5), including the correct value of the Rydberg constant R , in terms of known fundamental constants.

*The subscript γ is traditionally used to identify variables pertaining to any kind of photon, since "gamma" (γ) is one of the many names for a photon (although this is a little inconsistent since the name gamma is usually reserved for photons of very high energy).



5.4 The Problem of Atomic Stability

A satisfactory theory of atomic spectra obviously required a correct knowledge of the structure of the atom. It is therefore not surprising that atomic spectra lacked any explanation until Rutherford had proposed his nuclear model of the atom in 1911, nor that Bohr's explanation came soon after that proposal, in 1913. Curiously enough, the nuclear model posed a serious problem of its own, and it was in solving this problem that Bohr succeeded in explaining atomic spectra as well.

As Rutherford himself was aware, his model of the atom raises an awkward problem of stability. Superficially, Rutherford's atom resembles the solar system, and electrons can orbit around the nucleus just as planets orbit around the sun. According to classical mechanics, the planets' orbits are circles or ellipses and are *stable*: Once a planet is placed in a given orbit it will remain there indefinitely (if we ignore small effects like tidal friction). Unfortunately, the same is not true in an atom. A clear prediction of classical electromagnetic theory is that an accelerating charge should radiate electromagnetic waves. The electron has an electric charge and should have a centripetal acceleration in its "planetary" orbit. Therefore, the orbiting electron should radiate electromagnetic waves and hence gradually lose energy. This implies that its orbital radius should steadily shrink and its orbital frequency increase. The steady increase in orbital frequency means that the frequency of emitted light should keep changing — in sharp contrast to the observed spectrum, with its discrete, fixed frequencies. Worse still, one can estimate the rate at which the radius will shrink, and one finds that all electrons should collapse into the nucleus in a time of order 10^{-11} s. (See Section 11.2.) That is, stable atoms, as we know them, could not exist.* This was the problem that Bohr originally set out to solve.

5.5 Bohr's Explanation of Atomic Spectra

To solve the problem of atomic stability, Bohr proposed that the laws of classical mechanics must be modified and that among the continuum of electron orbits that classical mechanics predicts, only a certain discrete set is actually possible. He gave these allowed orbits the name **stationary orbits** or **stationary states**. Since the possible orbits were discrete, their energies would also be discrete; that is, the energies of the electrons in an atom would be quantized, and the only possible energies of the whole atom would be a discrete set E_1, E_2, E_3, \dots . If this were true, it would be impossible for the atom to lose energy steadily and continuously, as required by classical electromagnetic theory. Therefore, Bohr simply postulated that an electron in one of the allowed stationary orbits does not radiate energy and remains in exactly the same orbit as long as it is not disturbed.

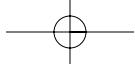
Bohr could not show *why* the electrons in his stationary states do not radiate, and one cannot really say that he *explained* the stability of atoms.

Niels Bohr
(1885–1962, Danish)



After getting his PhD at Copenhagen, Bohr worked under J. J. Thomson at Cambridge and then Rutherford at Manchester. It was during this period that he devised his model of the hydrogen atom (1913), which was the first reasonably successful explanation of atomic spectra in terms of the atom's internal structure and was an essential step on the path to modern quantum theory. It also earned him the 1922 Nobel Prize in physics. Bohr continued to be active in the development of modern physics — including the philosophical underpinnings of quantum theory — and set up an institute in Copenhagen that was one of the world's great centers for theoretical physics, both before and after World War II. It was Bohr who brought the news of nuclear fission to the United States in 1939 and set in motion the development of nuclear bombs.

*In theory, we should consider the same problem with respect to the solar system. As the planets move in their orbits, they should radiate gravitational waves, analogous to the electromagnetic waves radiated by electrons. However, gravitational radiation is so small that it has not yet (2002) been detected directly, and its effect on the planetary orbits is certainly unimportant.



150 Chapter 5 • Quantization of Atomic Energy Levels

Nevertheless, his ideas come very close to being correct, and his phrase “stationary state” has proved remarkably apt. In modern quantum mechanics we will find that the electron does not have a classical orbit at all; rather, it is (in a sense we’ll explain in Chapter 6) distributed continuously through the atom and can be visualized as a cloud of charge surrounding the nucleus. The stable states of the atom — corresponding to Bohr’s stationary orbits — are states in which the distribution of charge in this cloud *actually is stationary* and does not radiate.*

Having solved the problem of atomic stability (by simply postulating his quantized stationary states, in which electrons did not radiate), Bohr realized that his theory gave a beautiful explanation of atomic spectra. As we have already described, if the total energy of the electrons in an atom is quantized, with allowed values E_1, E_2, E_3, \dots , the energy can change only by making a discontinuous transition from one value E_n to another $E_{n'}$,

$$E_n \rightarrow E_{n'}$$

Bohr did not try to explain the detailed mechanisms by which such a transition can occur, but one method is certainly the emission or absorption of a photon. If $E_n > E_{n'}$, a photon of energy $E_n - E_{n'}$ must be *emitted*. If $E_n < E_{n'}$, a photon of energy $E_{n'} - E_n$ must be *absorbed*. Either way, the photon’s energy must be the difference of two of the allowed energies, E_n and $E_{n'}$ of the atom. This immediately explains why the energies (and hence frequencies) of the emitted and absorbed photons are the same. Further, one would naturally expect the allowed energies E_1, E_2, E_3, \dots to be different for different atomic species. Thus, the same should be true of the differences $E_n - E_{n'}$, and this would explain why each atomic species emits and absorbs with its own characteristic spectrum.

Example 5.2

The helium atom has two stationary states, designated $3p$ and $2s$, with energies

$$E_{3p} = 23.1 \text{ eV} \quad \text{and} \quad E_{2s} = 20.6 \text{ eV}$$

measured on a scale in which the lowest energy state has an energy set to zero. (We will see the significance of the designations $3p$ and $2s$ later.) What will be the wavelength of a photon emitted when the atom makes a transition from the $3p$ state to the $2s$?

In moving from the $3p$ state to the $2s$, the atom loses energy $E_{3p} - E_{2s}$. This is therefore the energy of the emitted photon

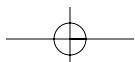
$$E_\gamma = E_{3p} - E_{2s} = 2.5 \text{ eV}$$

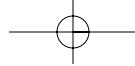
so the wavelength is

$$\lambda = \frac{hc}{E_\gamma} = \frac{1240 \text{ eV} \cdot \text{nm}}{2.5 \text{ eV}} \approx 500 \text{ nm}$$

Light with this wavelength is blue-green, and the $3p \rightarrow 2s$ transition is, in fact, responsible for the blue-green line that is visible in the helium spectrum inside the front cover.

* We will discuss this further in Section 7.3.





5.6 The Bohr Model of the Hydrogen Atom

It was obviously desirable that Bohr find a way to predict the allowed energies, E_1, E_2, \dots , of an atom, and in the case of hydrogen he was able to do so. In fact, he produced several different arguments, all of which gave the same answer for the allowed energies. All these arguments were, as Bohr himself acknowledged, quite tentative, and their main justification was that they produced the right answer; that is, they predicted certain energy levels, which in turn led to the Rydberg formula for the spectrum of hydrogen.

We will describe one of Bohr's arguments, which is the simplest and, in many ways, the closest to modern quantum mechanics. Since Bohr assumed that the possible orbits of the electron were a subset of the classical orbits, we begin by reviewing the classical mechanics of an orbiting electron.

Our system consists of an electron of mass m and charge $-e$, which orbits around a proton of charge $+e$, as shown in Fig. 5.3. For simplicity, we will treat the case where the electron moves in a circular orbit, and because the proton is so much heavier than the electron, we will make the approximation that the proton is fixed in position. (In reality, the proton moves a little, and this requires a very small correction to our answers, as we discuss later.)

The electron's acceleration is the centripetal acceleration, $a = v^2/r$, and the only force acting on the electron is the Coulomb attraction of the proton,

$$F = \frac{ke^2}{r^2}$$

where k is the Coulomb force constant, $k = 1/(4\pi\epsilon_0) = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$. Thus Newton's second law implies that

$$(\text{mass}) \times (\text{centripetal acceleration}) = \text{Coulomb force}$$

or

$$m \frac{v^2}{r} = \frac{ke^2}{r^2} \quad (5.6)$$

This condition is a relation between v and r , which can be solved to give v in terms of r or vice versa.

In classical mechanics (5.6) is the only constraint between v and r , so neither v nor r is fixed. On the contrary, the possible values of v and r range continuously from 0 to ∞ , and this means that the energy of the electron is not quantized. To see this explicitly, we note that (5.6) implies that

$$mv^2 = \frac{ke^2}{r} \quad (5.7)$$

Now, the electron's kinetic energy K is $K = \frac{1}{2}mv^2$, while the potential energy of an electron (charge $-e$) in the field of a proton (charge $+e$) is

$$U = -\frac{ke^2}{r} \quad (5.8)$$

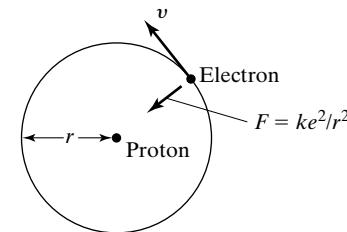
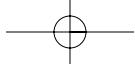


FIGURE 5.3

In the Bohr model, the hydrogen atom consists of an electron in orbit around a proton. The centripetal acceleration, $a = v^2/r$, is supplied by the Coulomb attraction, $F = ke^2/r^2$.



152 Chapter 5 • Quantization of Atomic Energy Levels

if we define U to be zero when the electron is far from the proton (that is, when $r = \infty$). Thus (5.7) implies that for an electron in a circular orbit

$$K = -\frac{1}{2}U \quad (5.9)$$

(This result was well known in classical mechanics and is an example of the so-called virial theorem.) The total energy is therefore

$$E = K + U = \frac{1}{2}U = -\frac{1}{2}\frac{ke^2}{r} \quad (5.10)$$

Notice that the total energy is negative, as it has to be since the electron is bound to the proton and cannot escape to infinity. Since r can have any value in the range $0 < r < \infty$, it is clear from (5.10) that the energy of our bound electron can have any value in the range $-\infty < E < 0$.

Our analysis so far has been purely classical. Some new hypothesis was needed if the allowed energies were to be quantized. To understand the hypothesis that Bohr proposed, we note that Planck's constant has the same dimensions as angular momentum (remember that since $E = hf$, h has the units of energy/frequency or energy \times time):

$$[h] = \text{energy} \times \text{time} = \frac{ML^2}{T^2} \times T = \frac{ML^2}{T}$$

and

$$[\text{angular momentum}] = [mvr] = M \times \frac{L}{T} \times L = \frac{ML^2}{T} \quad (5.11)$$

This suggests that the electron's angular momentum could be quantized in multiples of h ; and Bohr proposed specifically that its allowed values are integer multiples of $h/(2\pi)$:

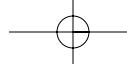
$$L = \frac{h}{2\pi}, \quad 2\frac{h}{2\pi}, \quad 3\frac{h}{2\pi}, \dots \quad (5.12)$$

where L denotes the electron's angular momentum. Bohr was led to propose these values for L by what he called the correspondence principle, which we will describe briefly in Problem 11.12. For the moment, we simply accept (5.12) as a judicious guess that, like Planck's hypothesis that light is quantized in multiples of hf , was principally justified by the fact that it led to the correct answers. When we go on to discuss modern quantum mechanics we will be able to *prove* that **Bohr's quantization condition** (5.12) is essentially correct.*

The combination $h/(2\pi)$ in (5.12) appears so frequently that it is often given its own symbol:

$$\hbar = \frac{h}{2\pi} = 1.054 \times 10^{-34} \text{ J} \cdot \text{s} \quad (5.13)$$

*The hypothesis (5.12) is not exactly correct. We will prove, rather, that any *component* of the vector \mathbf{L} is an integer multiple of $h/2\pi$. However, for the present this comes to the same thing because we can take the electron's orbit to lie in the xy plane, in which case the total angular momentum is the same as its z component.

Section 5.6 • The Bohr Model of the Hydrogen Atom **153**

where \hbar is read as “ h bar.” Thus, we can rewrite the quantization condition (5.12) as

$$L = n\hbar \quad (n = 1, 2, 3, \dots) \quad (5.14)$$

For the circular orbits, which we are discussing, the angular momentum is $L = mvr$, and (5.14) can be rewritten as

$$mvr = n\hbar \quad (n = 1, 2, 3, \dots) \quad (5.15)$$

This condition is a second relation between r and v . [The first was (5.7), which expressed Newton’s second law.] With two equations for two unknowns we can now solve to find the allowed values of r (or v). If we solve (5.15) to give $v = n\hbar/(mr)$ and then substitute into (5.7), we find

$$m\left(\frac{n\hbar}{mr}\right)^2 = \frac{ke^2}{r}$$

whence

$$r = \frac{n^2\hbar^2}{ke^2m} \quad (5.16)$$

That is, the values of r are quantized, with values given by (5.16), which we write as

$$r = n^2a_B \quad (n = 1, 2, 3, \dots) \quad (5.17)$$

Here we have defined the **Bohr radius** a_B , which is easily evaluated to be (Problem 5.7)

$$a_B = \frac{\hbar^2}{ke^2m} = 0.0529 \text{ nm} \quad (5.18)$$

Knowing the possible radii of the electron’s orbits, we can immediately find the possible energies from (5.10).

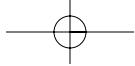
$$E = -\frac{ke^2}{2r} = -\frac{ke^2}{2a_B} \frac{1}{n^2} \quad (n = 1, 2, 3, \dots) \quad (5.19)$$

We see that the possible energies of the hydrogen atom are quantized. If we denote the energy (5.19) by E_n , then, as we argued in Section 5.5, the energy of a photon emitted or absorbed by hydrogen must have the form

$$E_\gamma = E_n - E_{n'} = \frac{ke^2}{2a_B} \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (5.20)$$

which has precisely the form of the Rydberg formula (5.5)

$$E_\gamma = hcR \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (5.21)$$



154 Chapter 5 • Quantization of Atomic Energy Levels

Comparing (5.20) and (5.21), we see that Bohr's theory predicts both the Rydberg formula and the value of the Rydberg constant,*

$$R = \frac{ke^2}{2a_B(hc)} = \frac{1.44 \text{ eV} \cdot \text{nm}}{2 \times (0.0529 \text{ nm}) \times (1240 \text{ eV} \cdot \text{nm})} \\ = 0.0110 \text{ nm}^{-1}$$

in perfect agreement with the observed value (5.2).

Because of its close connection with the Rydberg constant, the energy $hcR = ke^2/(2a_B)$ in (5.21) and (5.20) is called the **Rydberg energy** and is denoted E_R . Its value and several equivalent expressions for it are (as you should check for yourself in Problem 5.8)

$$E_R = hcR = \frac{ke^2}{2a_B} = \frac{m(ke^2)^2}{2\hbar^2} = 13.6 \text{ eV} \quad (5.22)$$

In terms of E_R , the allowed energies (5.19) of the electron in a hydrogen atom are

$$E_n = -\frac{E_R}{n^2} \quad (5.23)$$

This is the most important result of the Bohr model, and we take up its implications in the next section.

5.7 Properties of the Bohr Atom

As compared to the modern quantum-mechanical view, Bohr's model of the hydrogen atom is not completely correct. Nevertheless, it is correct in several important features and is often easier to visualize and remember than its modern counterpart. For these reasons we review some of its properties.

Bohr's model predicts (and modern quantum mechanics agrees) that the possible energies of an electron in a hydrogen atom are quantized, their allowed values being $E_n = -E_R/n^2$, where $n = 1, 2, \dots$. The lowest possible energy is that with $n = 1$ and is

$$E_1 = -E_R = -13.6 \text{ eV} \quad (5.24)$$

This state of lowest energy is called the **ground state**. It is the most stable state of the atom and is the state into which an isolated atom will eventually find its way. The significance of the energy $E_1 = -13.6 \text{ eV}$ is that an energy $+13.6 \text{ eV}$ must be supplied to remove the electron entirely from the proton. That is, the Bohr theory predicts that the binding energy of the hydrogen atom is 13.6 eV, in excellent agreement with its observed value.

According to (5.17) the radius of the $n = 1$ orbit is just the Bohr radius a_B :

$$r = a_B = 0.0529 \text{ nm} \quad (5.25)$$

*In evaluating this, we have used the two useful combinations $ke^2 = 1.44 \text{ eV} \cdot \text{nm}$ and $hc = 1240 \text{ eV} \cdot \text{nm}$. (See Problem 5.4.) These are listed, along with many other physical constants, inside the front cover.

Section 5.7 • Properties of the Bohr Atom **155**

This agrees well with the observed size of the hydrogen atom and was regarded by Bohr as an important accomplishment of his theory.* The primary significance of a_B is that it gives the radius of the ground state of hydrogen. However, we will find that it also gives the order of magnitude of the outer radius of all atoms in their ground states. For this reason, the Bohr radius a_B is often used as a unit of distance in atomic physics.

The orbits with energies greater than the ground-state energy are called *excited states*. Their energies are given by $E_n = -E_R/n^2$, with $n = 2, 3, \dots$ that is,

$$E_2 = -\frac{E_R}{4} = -3.4 \text{ eV} \quad E_3 = -\frac{E_R}{9} = -1.5 \text{ eV}$$

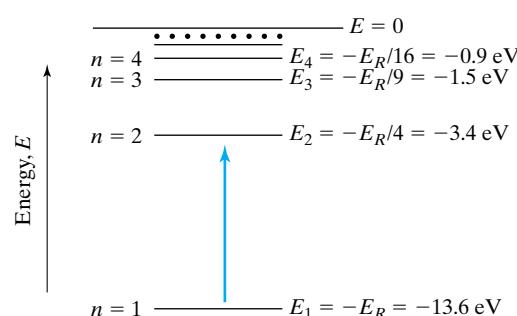
and so on. These allowed energies, or **energy levels**, are traditionally displayed graphically as in Fig. 5.4. In these **energy-level diagrams**, the energy is plotted vertically upward, and the allowed energies are shown as horizontal lines, somewhat like the rungs of a ladder.

Energy-level diagrams provide a convenient way to represent transitions between the energy levels. For example, if the atom is in the lowest state, $n = 1$, the only possible change is an upward transition, which will require the supply of some energy. If, for instance, we were to shine photons of energy 10.2 eV on a hydrogen atom, the photons would have exactly the right energy to lift the electron to the $n = 2$ level and the atom could make the transition, absorbing one photon in the process. This transition is indicated in Fig. 5.4 by an arrow between the levels concerned. The energy 10.2 eV is called the **first excitation energy** of hydrogen, since it is the energy required to raise the atom to its first excited state; similarly, the second excitation energy is $E_3 - E_1 = 12.1$ eV, and so on.

If the atom is in an excited state n (with $n > 1$), it can drop to a lower state n' (with $n' < n$) by emitting a photon of energy $E_n - E_{n'}$. If, for example, the original level is $n = 3, 4, 5, \dots$ and the electron drops to the $n = 2$ orbit, the photon will have energy

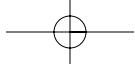
$$E_\gamma = E_n - E_2 = E_R \left(\frac{1}{2^2} - \frac{1}{n^2} \right) \quad (5.26)$$

These are, of course, the photon energies implied by Balmer's original formula (5.1). For this reason, the spectral lines given by this formula (with the lower

**FIGURE 5.4**

Energy levels of the hydrogen atom. The lowest level is called the ground state, and the higher levels, excited states. There are infinitely many levels, $n = 5, 6, 7, \dots$, all squeezed below $E = 0$. The upward arrow represents a transition in which the atom is excited from the ground state to the first excited level, by the absorption of an energy 10.2 eV. (Note that the diagram is not exactly to scale; the interval between E_1 and E_2 is really larger than shown.)

*We will see later that in modern quantum mechanics there is no uniquely defined radius of the electron's orbit. However, the average radius is about a_B ; thus in this sense, Bohr's theory agrees with modern quantum mechanics.



156 Chapter 5 • Quantization of Atomic Energy Levels

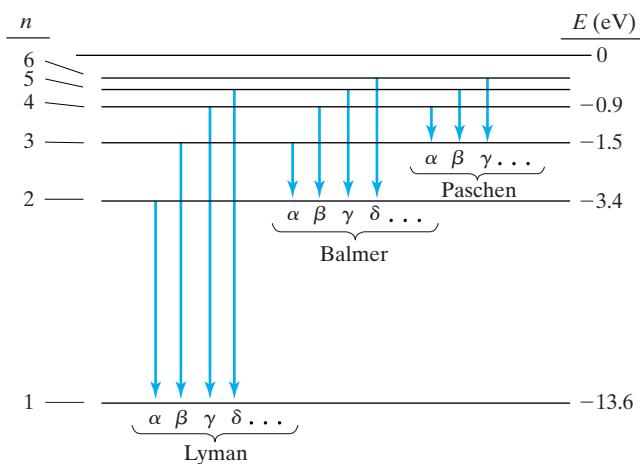


FIGURE 5.5

Some of the transitions of the Lyman, Balmer, and Paschen series in atomic hydrogen. The lines of each series are labeled α , β , γ , ..., starting with the line of longest wavelength, or least energy. In principle, each series has infinitely many lines. (Not to scale.)

level being $n' = 2$) are often called the **Balmer series**. The transitions in which the lower level is the ground state ($n' = 1$) are called the Lyman series, and those in which the lower level is $n' = 3$, the Paschen series (after their respective discoverers). These three series are illustrated in Fig. 5.5.

According to (5.17), the radius of the n th circular orbit is proportional to n^2 :

$$r = n^2 a_B$$

Thus the radii of the Bohr orbits increase rapidly with n , as indicated in Fig. 5.6. This agrees qualitatively with modern quantum theory and with experiment: In the excited states of hydrogen atoms (and all other atoms, in fact) the electrons tend to be much farther away from the nucleus than they are in their ground states.

In addition to the circular orbits that we have discussed, Bohr's theory also allowed certain elliptical orbits. However, Bohr was able to show that the allowed energies of these elliptical orbits were the same as those of the circular orbits. Thus, for our present purposes the elliptical orbits do not add any important further information. Since the precise details of the Bohr orbits are not correct anyway, we will not discuss the elliptical orbits any further.

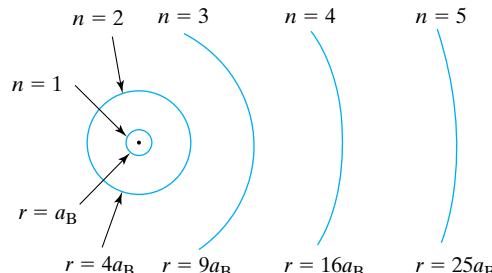


FIGURE 5.6

The radius of the n th Bohr orbit is $n^2 a_B$. (Drawn to scale.)

Example 5.3

Modern atomic physicists have observed hydrogen atoms in states with $n > 100$. What is the diameter of a hydrogen atom with $n = 100$?

The diameter is

$$d = 2r = 2n^2 a_B = 2 \times 10^4 \times (0.05 \text{ nm}) = 1 \mu\text{m}.$$

By atomic standards this is an enormous size — 10^4 times the diameter in the ground state. For comparison, we note that a quartz fiber with this diameter is visible to the naked eye. Atoms with these high values of n — called Rydberg atoms — can exist only in a good vacuum since the interatomic spacing at normal pressures is of order 3 nm and leaves no room for atoms this large. (See Problem 5.14.)

5.8 Hydrogen-Like Ions

For several years, at least, it appeared that Bohr's theory gave a perfect account of the hydrogen atom. The important problem was to generalize the theory to atoms with more than one electron, and in this no one succeeded. In fact, the Bohr theory was never successfully generalized to explain multielectron atoms, and a satisfactory quantitative theory had to await the development of modern quantum mechanics around 1925. Nevertheless, the Bohr theory did give a successful quantitative account of some atomic problems besides the spectrum of atomic hydrogen. In this and the next section we describe two of these successes.

In Section 5.6 we found the allowed radii and energies of a hydrogen atom, that is, a single electron in orbit around a proton of charge $+e$. The arguments given there started with Bohr's quantization condition that the allowed values of angular momentum are integer multiples of \hbar . Starting from this same assumption, we can modify those arguments to apply to any **hydrogen-like ion**, that is, any atom that has lost all but one of its electrons and therefore comprises a single electron in orbit around a nucleus of charge $+Ze$. We might consider, for example, the He^+ ion (an electron and a helium nucleus of charge $+2e$) or the Li^{2+} ion (an electron and a lithium nucleus of charge $+3e$).

To adapt the arguments of Section 5.6 to hydrogen-like ions, we have only to note that the force ke^2/r^2 on the electron in hydrogen must be replaced by

$$F = \frac{Zke^2}{r^2}$$

In other words, wherever ke^2 appears in Section 5.6, it must be replaced by Zke^2 . For example, the allowed orbits of hydrogen had radii given by (5.16) as

$$r = n^2 \frac{\hbar^2}{ke^2 m} = n^2 a_B$$

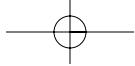
therefore, the orbits of an electron moving around a charge Ze are

$$r = n^2 \frac{\hbar^2}{Zke^2 m} = n^2 \frac{a_B}{Z} \quad (5.27)$$

We see that the radius of any given orbit is inversely proportional to Z . The larger the nuclear charge Z , the closer the electron is pulled in toward the nucleus, just as one might expect.

The potential energy of the hydrogen-like ion is $U = -Zke^2/r$. The total energy is, according to (5.10), $E = K + U = U/2$ or

$$E = -\frac{Zke^2}{2r} \quad (5.28)$$



158 Chapter 5 • Quantization of Atomic Energy Levels

Inserting (5.27) for the radius of the n th orbit, we find that

$$E_n = -Z^2 \frac{ke^2}{2a_B} \frac{1}{n^2}$$

or

$$E_n = -Z^2 \frac{E_R}{n^2} \quad (5.29)$$

That is, the allowed energies of the hydrogen-like ion with nuclear charge Ze are Z^2 times the corresponding energies in hydrogen. [The two factors of Z are easy to understand: One is the Z in the expression (5.28) for the energy, the other comes from the $1/Z$ in the allowed radii.]

The result (5.29) implies that the energy levels of the He^+ ion should be four times those in hydrogen. Thus the energies of the photons emitted and absorbed by He^+ should be

$$E_\gamma = 4E_R \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (5.30)$$

(that is, four times those of the hydrogen atom). This formula looks so like the Rydberg formula for hydrogen that when the spectrum of He^+ had been observed in 1896 in light from the star Zeta Puppis, it had been wrongly interpreted as a new series of lines for hydrogen. It was another of the triumphs for Bohr's theory that he could explain these lines as belonging to the spectrum of once-ionized helium. Today, the spectra of hydrogen-like ions ranging from He^+ and Li^{2+} to Fe^{25+} (iron with 25 of its 26 electrons removed) have been observed and are in excellent agreement with the Bohr formula (5.29).

There is a small but interesting correction to (5.29) that we should mention. So far we have supposed that our one electron orbits around a fixed nucleus; in reality, the electron and nucleus both orbit around their common center of mass. Because the electron is so light compared to the nucleus, the center of mass is very close to the nucleus. Thus the nucleus is very nearly stationary, and our approximation is very good. Nonetheless, the nucleus *does* move, and this motion is fairly easily taken into account. In particular, it can be shown (Problem 5.21) that the allowed energies are still $E = -Z^2 E_R / n^2$ and that the Rydberg energy is still given by (5.22) as

$$E_R = \frac{m(ke^2)^2}{2\hbar^2} \quad (5.31)$$

provided that the mass m of the electron is replaced by the so-called **reduced mass**, usually denoted μ :

$$\mu = \text{reduced mass} = \frac{m}{1 + m/m_{\text{nuc}}} \quad (5.32)$$

where m is the electron mass and m_{nuc} , the mass of the nucleus. (See Problem 5.21.) In the case of hydrogen, the nucleus is a proton and $m/m_{\text{nuc}} \approx 1/1800$. Therefore, the energy levels in hydrogen are all reduced by about 1 part in 1800 as a result of this correction. This is a rather small change, but one that can be easily detected by the careful spectroscopist.

The interesting thing (for the present discussion) is that the correction represented by (5.32) is different for different nuclei. In He^+ the nucleus is four times heavier than in hydrogen, and the factor m/m_{nuc} is four times smaller. Because of this difference, the ratio of the He^+ frequencies to those of hydrogen is not exactly 4 but is about 4.002. This small difference was observed and added further weight to Bohr's interpretation.

Example 5.4

Hydrogen has an isotope, ${}^2\text{H}$, called deuterium or heavy hydrogen, whose nucleus is almost exactly twice as heavy as that of ordinary hydrogen since it contains a proton *and* a neutron. It was discovered because its spectrum is not exactly the same as that of ordinary hydrogen. Calculate the Balmer α wavelengths of ordinary hydrogen and of deuterium, both to five significant figures, and compare.

Since ${}^1\text{H}$ and ${}^2\text{H}$ have the same nuclear charge, they would have identical spectra, if it were not for the motion of the nuclei. In particular, the Balmer α line would be given by the Rydberg formula (5.4) with $n = 3$, $n' = 2$, and $R = 0.0109737 \text{ nm}^{-1}$ (to six significant figures); this would give

$$\lambda = 656.114 \text{ nm}$$

However, all energy levels of each atom must be corrected in accordance with (5.32) and (5.31) by dividing by the factor $(1 + m/m_{\text{nuc}})$. Since wavelength is inversely proportional to energy, the correct wavelengths are found by multiplying by this same factor. Therefore, for ordinary hydrogen, the Balmer α wavelength is really

$$\begin{aligned}\lambda({}^1\text{H}) &= (656.114 \text{ nm}) \left(1 + \frac{m}{m({}^1\text{H})} \right) \\ &= (656.114 \text{ nm}) \left(1 + \frac{1}{1800} \right) = 656.48 \text{ nm}\end{aligned}$$

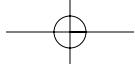
For deuterium the corresponding wavelength is

$$\lambda({}^2\text{H}) = (656.114 \text{ nm}) \left(1 + \frac{1}{3600} \right) = 656.30 \text{ nm}$$

a difference of about 1 part in 4000.

Since natural hydrogen contains 0.015% deuterium, its spectrum has a very faint component with these slightly shorter wavelengths. It was by observing these lines that the American chemist Harold Urey proved the existence of deuterium in 1931.

The reduced mass (5.32) is always less than the mass of the electron, and the difference is bigger for lighter nuclei. The extreme case is the system called positronium, which is a hydrogen-like “atom” with the electron bound to a positron. (Positrons were described in Section 2.8.) In this case the “nuclear” mass is equal to the electron mass, and the reduced mass is exactly half the electron mass. Thus the energy levels of positronium are just half those of hydrogen. (See Problem 5.18.) The other interesting thing about positronium is that it is very unstable since the electron and positron eventually annihilate, producing a 1.22 MeV of electromagnetic energy.



160 Chapter 5 • Quantization of Atomic Energy Levels

5.9 X-ray Spectra

The quantitative successes of Bohr's theory all concerned systems in which a single electron moves in the field of a single positive charge. The most obvious example is the hydrogen-like ion discussed in Section 5.8, but another system that fits the description, at least approximately, is the innermost electron of a multielectron atom. To the extent that the charge distribution of the other outer electrons is spherical (which is actually true to a fair approximation), the outer electrons exert no net force on the innermost electron.* Therefore, the latter feels only the force of the nuclear charge Ze , and its allowed energies should be given by (5.29) as about

$$E_n = -Z^2 \frac{E_R}{n^2} \quad (5.33)$$

The factor Z^2 means that for medium and heavy atoms the inner electron is very tightly bound. For example, in zinc, with $Z = 30$, the energy needed to remove the innermost electron from the $n = 1$ orbit is about

$$Z^2 E_R = (30)^2 \times (13.6 \text{ eV}) \approx 12,000 \text{ eV}$$

Thus, atomic transitions that involve the inner electrons would be expected to involve energies of order several thousand eV; in particular, a photon emitted or absorbed in such a transition should be an X-ray photon. This fact was recognized by the young British physicist Henry Moseley, who, within a few months of Bohr's paper, had shown that the Bohr theory gave a beautiful explanation of the characteristic X-rays that were produced at discrete frequencies, characteristic of the anode material of an X-ray tube, as discussed in Section 4.5.

Moseley's explanation of the characteristic X-rays was very simple: In an X-ray tube, the anode is struck by high-energy electrons, which can eject one or more of the electrons in the anode. If an electron in the $n = 1$ orbit is ejected, this will create a vacancy in the $n = 1$ level, into which an outer atomic electron can now fall.[†] If, for example, an $n = 2$ electron falls into this $n = 1$ vacancy, a photon will be emitted with energy given by (5.33) as

$$E_\gamma = E_2 - E_1 = Z^2 E_R \left(1 - \frac{1}{4}\right) = \frac{3}{4} Z^2 E_R \quad (5.34)$$

Transitions between $n = 2$ and $n = 1$ are traditionally identified as the K_α transitions. Using this terminology, we can say that the Bohr theory predicts the K_α photons emitted or absorbed by an atom should have energy $E_\gamma = 3Z^2 E_R/4$. If one observes several different elements and measures the

Henry Moseley
(1887–1915, English)



Moseley exploited the newly discovered X-ray diffraction to measure the wavelengths of X-rays emitted by atoms. He showed that the wavelengths depend on the atomic number exactly as predicted by the Bohr model, and he used this dependence to identify unambiguously several previously uncertain atomic numbers. He was killed at age 27 in World War I.

*Remember that the field due to a spherically symmetric shell of charge is zero inside the shell.

[†]Implicit in this argument is the idea that each orbit can hold only a certain number of electrons. We will see in Chapter 10 that this is true, because of the so-called Pauli exclusion principle.

frequencies of their K_{α} X-rays (or any other definite X-ray line), then the photon energies, and hence frequencies, should vary like the square of the atomic number Z ; that is, we should find $f \propto Z^2$, or equivalently

$$\sqrt{f} \propto Z \quad (5.35)$$

Moseley measured the K_{α} lines of some 20 elements. By plotting \sqrt{f} against the known values of Z and showing that the data fitted a straight line (Fig. 5.7), he verified the prediction (5.35) and gave strong support to the Bohr theory. At that time (1913) the significance of the atomic number Z as the number of positive charges on the nucleus was only just becoming apparent, and Moseley's work settled this point conclusively. The atomic numbers of several elements were still in doubt, and Moseley's data, plotted as in Fig. 5.7, allowed these numbers to be determined unambiguously. Moseley was also able to identify three atomic numbers for which the corresponding elements had not yet been found — for example, $Z = 43$, technetium, which does not occur naturally and was first produced artificially in 1937.

A close look at Fig. 5.7 shows that the data do not confirm the prediction (5.35) exactly. If $\sqrt{f} \propto Z$, the line in Fig. 5.7 should pass through the origin, which it does not quite do. The line shown (which is a least-squares best fit) meets the Z axis close to $Z = 1$. That is, the data show that $\sqrt{f} \propto (Z - 1)$, or equivalently,

$$E_{\gamma} \propto (Z - 1)^2 \quad (5.36)$$

This small discrepancy was explained (and, in fact, anticipated) by Moseley, as follows: The prediction that the X-ray frequencies of a given line should be proportional to Z^2 was based on the assumption that the inner electron feels only the force of the nuclear charge Ze and is completely unaffected by any of the other electrons. This is a fair approximation, but certainly not perfect. An inner electron does experience some repulsion by the other electrons, and this slightly offsets, or *screens*, the attraction of the nucleus. This amounts to a small reduction in the nuclear charge, which we can represent by replacing Ze with $(Z - \delta)e$, where δ is some (unknown) small number. In this case the energy levels of the inner electron, and hence the X-ray energies, should be proportional to $(Z - \delta)^2$ rather than Z^2 ; specifically, (5.34) should be replaced by

$$E_{\gamma} = \frac{3}{4}(Z - \delta)^2 E_R \quad (5.37)$$

According to (5.36), the observed data fit this prediction perfectly, with a screening factor δ close to $\delta = 1$.

As we have seen, the dependence of the characteristic X-ray frequencies on atomic number is very simple (namely, f approximately proportional to Z^2). Also, because the transitions involve the inner electrons, the frequencies are independent of the external conditions of the atom (for example, whether it is bound to other atoms in a molecule or a solid). Further, Moseley found that with an impure anode he could easily detect the X-ray lines of the impurities. For all these reasons, he predicted that X-ray spectroscopy would "prove a powerful method of chemical analysis." This prediction has proved correct. In modern X-ray spectroscopy, a sample (a biological tissue, for example) is put in a beam of electrons, protons, or X-rays. The beam ejects inner electrons of many of the atoms in the sample, which then emit X-rays. By measuring the wavelengths emitted, one can identify all elements in the sample, down to the "trace" level of one part per million, or even less.

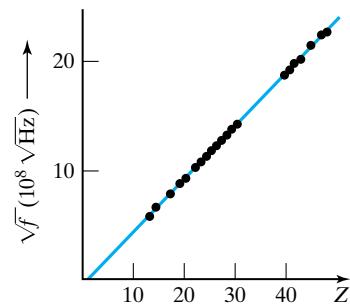
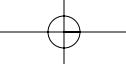


FIGURE 5.7

Moseley measured the frequencies f of K_{α} X-rays, using several different elements for the anode of his X-ray tube. The graph shows clearly that \sqrt{f} is a linear function of the atomic number Z of the anode material. The reason the line crosses the Z axis at $Z \approx 1$ is explained in the text.



162 Chapter 5 • Quantization of Atomic Energy Levels

Example 5.5

Most X-ray spectrometers have a thin window through which the X-rays must pass. Although high-energy X-rays pass easily through such windows, low-energy X-rays are severely attenuated and cannot be analyzed. A certain spectrometer, which is used for chemical analysis, cannot detect X-rays with $E_\gamma \leq 2.4 \text{ keV}$. If an unambiguous identification requires that one observe the K_α line of an element, what is the lightest element that can be identified using this spectrometer?

The energy of a K_α photon emitted by an element Z is given by (5.37) as $E_\gamma = 3(Z - \delta)^2 E_R / 4$ (with $\delta \approx 1$). The lowest detectable element is found by equating this energy to 2.4 keV and solving for Z , to give

$$Z = \delta + \sqrt{\frac{4E_\gamma}{3E_R}} = 1 + \sqrt{\frac{4 \times (2.4 \times 10^3 \text{ eV})}{3 \times (13.6 \text{ eV})}} = 16.3$$

We see from the table inside the back cover that the element with $Z = 17$ is chlorine and that with $Z = 16$ is sulfur. Therefore, our spectrometer can detect chlorine, but cannot detect sulfur.

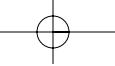
5.10 Other Evidence for Atomic Energy Levels*

* This section can be omitted without loss of continuity.

Although atomic spectroscopy gives abundant evidence for the quantization of atomic energy levels, one might hope to find other types of evidence as well. And, in fact, almost any process that transfers energy to or from an atom provides us with such evidence.

Imagine, for example, we fire a stream of electrons, all with the same kinetic energy K_0 , at a target of stationary atoms; and suppose, for simplicity, that all the atoms are in their ground state, with energy E_1 . The possible collisions between any one electron and an atom can be divided into two classes, the elastic and the inelastic. An **elastic collision** is defined as one in which the atom's internal state of motion is unaltered; this means that the total kinetic energy (of the incident electron plus the atom) does not change. Since the atom can recoil as a whole, it can gain some kinetic energy; but because the atom is so heavy compared to the electron, this recoil energy is very small (Problem 5.27). Therefore, for most purposes, an elastic collision can be characterized as one in which the scattered electron is deflected by the atom but suffers no appreciable loss of kinetic energy.

An **inelastic collision** is one in which the atom is excited to a different energy level and there is a corresponding reduction of the electron's kinetic energy. Because the atomic energy levels are quantized, the same must be true of the energy lost by the electron. Specifically, the electron can lose kinetic energy only in amounts equal to $E_n - E_1$, where E_n is an allowed energy of the atom. In particular, if the original kinetic energy K_0 is less than the first excitation energy, $E_2 - E_1$ of the atom, the electrons cannot excite the atoms at all and only elastic scattering is possible. Obviously, by studying electron scattering at various incident energies and by measuring how much energy the electrons lose, one should be able to demonstrate and measure the allowed energies of the atoms.



The Franck–Hertz Experiment

The first experiment along these lines was carried out by the German physicists James Franck and Gustav Hertz in 1914 and is duplicated in many undergraduate teaching laboratories today. In this experiment a stream of electrons is passed through a tube of mercury vapor, as shown schematically in Fig. 5.8(a). The electrons leave a heated cathode and are attracted toward the grid by an adjustable accelerating potential V_0 . Those electrons that pass through the grid will reach the anode provided that they have enough energy to overcome the small retarding potential ΔV . The current i reaching the anode is measured, and the observed behavior of i as a function of the accelerating potential V_0 is shown in Fig. 5.8(b). This behavior is easily explained in terms of the quantized energy levels of the mercury atom; in particular, the abrupt drop in the current each time the accelerating potential reaches a multiple of 4.9 V shows that the first excitation energy of mercury is 4.9 eV, as we now argue.

Even when the accelerating potential V_0 is zero, the heated cathode emits some electrons, but these collect in a cloud around the cathode, and the resulting field prevents the emission of any more electrons; so no steady current flows. When V_0 is slowly increased, some of the electrons in this cloud are drawn away, which allows more electrons to be emitted, and a current now flows. When an electron is accelerated by a potential difference V_0 , it acquires an energy V_0e . As long as this energy is less than the first excitation energy of the mercury atom, only elastic collisions are possible, and there is no way for the electrons to lose any energy to the vapor. (See Problem 5.27.) Thus, until V_0e reaches the first excitation energy, the anode current increases steadily as we increase V_0 . Once V_0e reaches the excitation energy, some electrons can excite the mercury atoms and lose most of their energy as a result. When these electrons reach the grid, they have insufficient energy to overcome the retarding potential ΔV and cannot reach the anode. Thus when V_0e reaches the first excitation energy, we expect a drop in the current i . The fact that the drop is observed when $V_0 = 4.9$ eV shows that the first excitation energy of mercury is 4.9 eV.

When V_0 increases beyond 4.9 V, the current increases again. But when V_0e reaches twice the excitation energy, some electrons can undergo two inelastic collisions between the cathode and grid, and the current drops again. This process continues, and it is possible under favorable conditions to observe 10 or more drops in the current, regularly spaced at intervals of 4.9 V.*

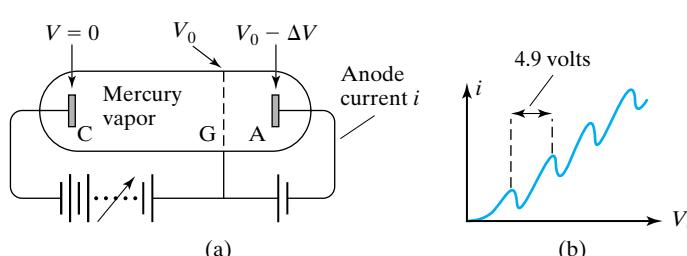
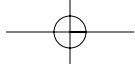


FIGURE 5.8

The Franck–Hertz experiment.
(a) Electrons leave the heated cathode C and pass through mercury vapor to the anode A. The grid G is kept at a higher potential, V_0 , than C and attracts the electrons; the anode is kept at a slightly lower potential $V_0 - \Delta V$.
(b) The anode current i as a function of accelerating potential V_0 .

*One can sometimes observe drops in i at voltages corresponding to the excitation of higher levels as well. However, the experiment is usually arranged in such a way that the probability for exciting the first excited state is very large and, hence, that very few electrons acquire enough energy to excite any higher levels.



164 Chapter 5 • Quantization of Atomic Energy Levels

This interpretation of the Franck–Hertz experiment can be confirmed by examining the optical spectrum of mercury. If 4.9 eV is the first excitation energy, the mercury atom should be able to emit and absorb photons of energy 4.9 eV. That is, the spectrum of mercury should include a line whose wavelength is

$$\lambda = \frac{hc}{E_\gamma} = \frac{1240 \text{ eV} \cdot \text{nm}}{4.9 \text{ eV}} = 250 \text{ nm}$$

and this is, indeed, the wavelength of a prominent line in the spectrum of mercury. Better still, one finds that the mercury vapor in the Franck–Hertz experiment begins emitting this line just as soon as V_0 passes 4.9 V and excitation of the first excited state becomes possible.

Energy-Loss Spectra

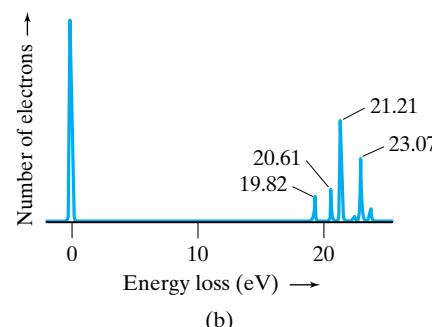
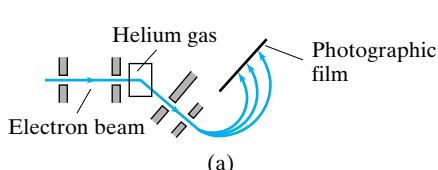
Today, energy levels of atoms (and, even more, nuclei) are routinely measured by finding the energy lost by inelastically scattered particles. Figure 5.9(a) is a schematic diagram of an arrangement for measuring the energy levels of the helium atom using inelastically scattered electrons. A beam of electrons, all with the same incident kinetic energy, is fired through a container of helium gas. Those electrons scattered at some convenient angle are sent through a magnetic field, which bends them into circular paths whose radii depend on their energies. [See Equation (2.48).] Therefore, a photographic film placed as shown lets one determine how many electrons are scattered at each different energy.

Figure 5.9(b) is a schematic plot of the number of scattered electrons as a function of their energy lost in the collision. This kind of plot is called an “energy-loss spectrum”—a natural generalization of the word “spectrum,” which in the context of light, this refers to numbers of photons as a function of their energy. The large peak on the left, which occurs at zero energy loss, corresponds to the many electrons that scatter elastically and hence lose no energy. The next peak occurs at an energy loss of 19.8 eV, indicating that the first excited state of helium is 19.8 eV above the ground state. The subsequent peaks indicate further excited states at 20.6 eV, 21.2 eV, and so on. All of the energies shown in Fig. 5.9 agree with the energy levels deduced from the optical spectrum, within the accuracy of the measurements.

Experiments like the Franck–Hertz experiment and the energy-loss measurements just described add confirmation (if any is needed) to Bohr’s hypothesis that atomic energy levels are quantized. In particular, they demonstrate clearly that the quantized energies of atomic spectra are more than just a property of the light emitted and absorbed by atoms; rather, as suggested by Bohr, they reflect the quantization of the atomic energy levels themselves.

FIGURE 5.9

Measurement of the energy levels of helium by inelastic scattering of electrons. (a) Electrons are scattered in helium gas. Those electrons scattered at one convenient angle are passed into a magnetic field, which bends them onto different paths according to their energy. (b) A typical energy-loss spectrum, showing the numbers of scattered electrons as a function of the energy that they lost in the collision. Each spike corresponds to the excitation of a particular atomic energy level. The different heights of the spikes depend on how easily the different levels can be excited. (Based on data of Trajmar, Rice, and Kuppermann taken at 25° with incident energy 34 eV.)



CHECKLIST FOR CHAPTER 5

Concept	Details
Atomic emission and absorption spectra	Light emitted and absorbed by the atom (Sec. 5.2) $E_\gamma = E_n' - E_n$
Balmer–Rydberg formula	Wavelengths of hydrogen's atomic spectrum $\frac{1}{\lambda} = R \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (5.4)$
Bohr model of hydrogen	Quantization of angular momentum, $L = n\hbar$ (5.14) Allowed radii, $r = n^2 a_B$ (5.17) Allowed energies, $E_n = -E_R/n^2$ (5.23) $a_B = \hbar^2/ke^2m = 0.0529 \text{ nm}$ (5.18) $E_R = ke^2/2a_B = 13.6 \text{ eV}$ (5.22)
Bohr radius	Picture of allowed energies, plotted upward
Rydberg energy	One electron bound to a charge Ze (Sec. 5.8) Allowed radii, $r = n^2 a_B/Z$ (5.27) Allowed energies, $E_n = -Z^2 E_R/n^2$ (5.29)
Energy-level diagrams	
Hydrogen-like ions	
X-ray spectra of different elements	$\sqrt{f} \propto Z$ (5.35)
Other evidence of atomic energy levels *	Franck–Hertz experiment Energy-loss spectra (Sec. 5.10)

PROBLEMS FOR CHAPTER 5

SECTION 5.3 (The Balmer–Rydberg Formula)

- 5.1** • Find the wavelength of the light emitted by hydrogen, as predicted by the Rydberg formula (5.4) with $n = 4$ and $n' = 3$. What is the nature of this radiation? (Visible? X-ray? etc.)
- 5.2** • Find the wavelength of the light emitted by hydrogen as predicted by the Rydberg formula (5.4) with $n = 4$ and $n' = 1$. What is the nature of this radiation? (Visible? X-ray? etc.)
- 5.3** • Use Equation (5.5) to calculate the upper limit of the energies of photons that can be emitted by a hydrogen atom (in eV). (This energy is called the Rydberg energy, as discussed in Section 5.6.) What is the lower limit?
- 5.4** • Starting from the SI values of k , e , h , and c , find the values of ke^2 and hc in $\text{eV} \cdot \text{nm}$. Prove, in particular, that

$$ke^2 = 1.44 \text{ eV} \cdot \text{nm}$$

- 5.5** •• Using the Rydberg formula (5.4), classify all of the spectral lines of atomic hydrogen as UV, visible, or IR.
- 5.6** •• The spectral lines of atomic hydrogen are given by the Rydberg formula (5.4). Those lines for which $n' = 1$ are called the Lyman series. Since n can be any integer greater than 1, there are (in principle, at least) infinitely many lines in the Lyman series. **(a)** Calculate the five longest wavelengths of the Lyman series. Mark the positions of these five lines along a linear scale of wavelength. **(b)** Prove that the

successive lines in the Lyman series get closer and closer together, approaching a definite limit (the *series limit*) as $n \rightarrow \infty$. Show this limit on your plot. What kind of radiation is the Lyman series? (Visible? X-ray? etc.)

SECTION 5.6 (The Bohr Model of the Hydrogen Atom)

- 5.7** • **(a)** Find the value of the Bohr radius $a_B = \hbar^2/(ke^2m)$ (where m is the electron's mass) by substituting the SI values of the constants concerned. **(b)** It is usually easier to do such calculations by using common combinations of constants, which can be memorized in convenient units ($ke^2 = 1.44 \text{ eV} \cdot \text{nm}$, for example). Find the value of the convenient combination $\hbar c$ in $\text{eV} \cdot \text{nm}$ from your knowledge of hc . [The value of hc was given in equation 4.8. Both hc and $\hbar c$ are worth remembering in $\text{eV} \cdot \text{nm}$.] Now calculate a_B by writing it as $(\hbar c)^2/(ke^2mc^2)$ and using known values of $\hbar c$, ke^2 , and mc^2 .

- 5.8** • Two equivalent definitions of the Rydberg energy E_R are

$$E_R = \frac{ke^2}{2a_B} = \frac{m(ke^2)^2}{2\hbar^2}$$

- (a)** Using the definition (5.18) of a_B , verify that these two definitions are equivalent. **(b)** Find the value of E_R from each of these expressions. (In the first case use $ke^2 = 1.44 \text{ eV} \cdot \text{nm}$ and the known value of a_B ; in the second, multiply top and bottom by c^2 , and then use the known values of mc^2 , ke^2 , and $\hbar c$.)

166 Chapter 5 • Quantization of Atomic Energy Levels

- 5.9** • Consider a charge q_1 with mass m in a circular orbit around a fixed charge q_2 , with q_1 and q_2 of opposite sign. Show that the kinetic energy K is $-1/2$ times the potential energy U and hence that $E = K + U = U/2$. [Your arguments can parallel those leading to (5.10). The point is for you to make sure you understand those arguments and to check that the conclusion is true for any two charges of opposite sign.]

- 5.10** •• (a) Derive an expression for the electron's speed in the n th Bohr orbit. (b) Prove that the orbit with highest speed is the $n = 1$ orbit, with $v_1 = ke^2/\hbar$. Compare this with the speed of light, and comment on the validity of ignoring relativity (as we did) in discussing the hydrogen atom. (c) The ratio

$$\alpha = \frac{v_1}{c} = \frac{ke^2}{\hbar^2} \quad (5.38)$$

is called the "fine-structure constant" (for reasons that are discussed in Problems 9.22 and 9.23) and is generally quoted as $\alpha \approx 1/137$. Verify this value.

SECTION 5.7 (Properties of the Bohr Atom)

- 5.11** • Find the range of wavelengths in the Balmer series of hydrogen. Does the Balmer series lie completely in the visible region of the spectrum? If not, what other regions does it include?

- 5.12** •• Find the range of wavelengths in each of the Lyman, Balmer, and Paschen series of hydrogen. Show that the lines in the Lyman series are all in the UV region, those of the Paschen are all in the IR region, while the Balmer series is in the visible and UV regions. (Note that visible light ranges from violet at about 400 nm to deep red at about 700 nm.) Show that these three series do not overlap one another but that the next series, in which the lower level is $n = 4$, overlaps the Paschen series.

- 5.13** •• The negative muon is a subatomic particle with the same charge as the electron but a mass that is about 207 times greater: $m_\mu \approx 207 m_e$. A muon can be captured by a proton to form a "muonic hydrogen atom," with energy and radius given by the Bohr model, except that m_e must be replaced by m_μ . (a) What are the radius and energy of the first Bohr orbit in a muonic hydrogen atom? (b) What is the wavelength of the Lyman α line in muonic hydrogen? What sort of electromagnetic radiation is this? (Visible? IR? etc.) Treat the proton as fixed (although this is not such a good approximation here — see Problem 5.17).

- 5.14** ••• The average distance D between the atoms or molecules in a gas is of order $D \approx 3 \text{ nm}$ at atmospheric pressure and room temperature. This distance is much larger than typical atomic sizes, and it is, therefore, reasonable to treat an atom as an isolated system, as we did in our discussion of hydrogen. However, the Bohr theory predicts that the radius of the n th orbit is $n^2 a_B$. Thus for sufficiently large n , the atoms would be larger than the spaces between them and our simple theory would surely not apply. Therefore, one would not expect to observe energy levels

for which the atomic diameter $2n^2 a_B$ is of order D or more. (a) At normal densities (with $D \approx 3 \text{ nm}$) what is the largest n that you would expect to observe? (b) If one reduced the pressure to 1/1000 of atmospheric, what would be the largest n ? [Remember that the spacing D is proportional to $(\text{pressure})^{-1/3}$ for constant temperature.] (c) Modern experiments have found hydrogen atoms in levels with $n \approx 100$. What must be the pressure in these experiments?

SECTION 5.8 (Hydrogen-Like Ions)

- 5.15** • What are the energy and wavelength of photons in the Lyman α line of Fe^{2+} (an iron nucleus with all but one of its 26 electrons removed)? What kind of electromagnetic radiation is this? (Visible? UV? etc.)

- 5.16** • What is the radius of the $n = 1$ orbit in the O^{7+} ion? What are the wavelength and energy of photons in the Lyman α line of O^{7+} ?

- 5.17** • In most of this chapter we treated the atomic nucleus as fixed. This approximation (often a very good one) can be avoided by using the reduced mass, as described in connection with Equations (5.31) and (5.32). (a) What percent error do we make in the energy levels of ordinary hydrogen when we treat the proton as fixed? (b) Answer the same question for muonic hydrogen, which is a negative muon bound to a proton. (See Problem 5.13.)

- 5.18** • What is the ground-state energy of positronium, the bound state of an electron and a positron?

[HINT: Here you *must* allow for the motion of the "nucleus" (that is, the positron) by using the reduced mass, as described in connection with Equations (5.31) and (5.32).]

- 5.19** •• When the spectrum of once-ionized helium, He^+ , was first observed, it was interpreted as a newly discovered part of the hydrogen spectrum. The following two questions illustrate this confusion: (a) Show that alternate lines in the Balmer series of He^+ — that is, those lines given by the Rydberg formula (5.30) with the lower level $n' = 2$ — coincide with the lines of the Lyman series of hydrogen. (b) Show that all lines of He^+ could be interpreted (incorrectly) as belonging to hydrogen if one supposed that the numbers n and n' in the Rydberg formula for hydrogen could be *half-integers* as well as integers.

- 5.20** •• The negative pion, π^- , is a subatomic particle with the same charge as the electron but $m_\pi = 273 m_e$. A π^- can be captured into Bohr orbits around an atomic nucleus, with radius given by the Bohr formula (5.27), except that m_e must be replaced by m_π . (a) What is the orbital radius for a π^- captured in the $n = 1$ orbit by a carbon nucleus? (b) Given that the carbon nucleus has radius $R \approx 3 \times 10^{-15} \text{ m}$, can this orbit be formed? (c) Repeat parts (a) and (b) for a lead nucleus (nuclear radius $\approx 7 \times 10^{-15} \text{ m}$). (The required atomic numbers can be found in Appendix C.)

- 5.21** ••• In Section 5.6 we treated the H atom as if the electron moves around a fixed proton. In reality, both the electron and proton orbit around their center of

mass as shown in Fig. 5.10. Using this figure, you can repeat the analysis of Section 5.6, including the small effects of the proton's motion, as follows:

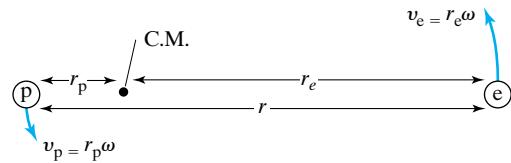


FIGURE 5.10

(Problem 5.21)

- (a) Write down the distances r_e and r_p in terms of r , m_e , and m_p .
- (b) Because both e and p move, it is easiest to work with the angular velocity ω , in terms of which v_e and v_p are as given in Fig. 5.10. Write down the total kinetic energy $K = K_e + K_p$ and prove that

$$K = \frac{1}{2}\mu r^2 \omega^2, \quad (5.39)$$

where μ is the *reduced mass*,

$$\mu = \frac{m_p m_e}{m_p + m_e} = \frac{m_e}{1 + (m_e/m_p)} \approx 0.9995 m_e \quad (5.40)$$

Notice that the expression (5.39) for K differs from its fixed proton counterpart, $K = (1/2)m_e r^2 \omega^2$, only in the replacement of m_e by μ .

- (c) Show that Newton's law, $F = ma$, applied to either the electron or proton gives

$$\frac{ke^2}{r^2} = \mu \omega^2 r \quad (5.41)$$

(Again, this differs from the fixed proton equivalent only in that μ has replaced m_e .)

- (d) Use (5.39) and (5.41) to show that $K = -U/2$ and $E = U/2$.
- (e) Show that the total angular momentum is

$$L = L_e + L_p = \mu r^2 \omega \quad (5.42)$$

(in place of $L = m_e r^2 \omega$ if the proton is fixed).

- (f) Assuming that the allowed values of L are $L = n\hbar$, where $n = 1, 2, 3, \dots$, use (5.41) and (5.42) to find the allowed radii r , and prove that the allowed energies are given by the usual formula $E = -E_R/n^2$, except that

$$E_R = \frac{\mu (ke^2)^2}{2\hbar^2} \quad (5.43)$$

This is the result quoted without proof in (5.31) and (5.32).

- (g) Calculate the energy of the ground state of hydrogen using (5.43), and compare with the result of using the fixed-proton result $E_R = m_e (ke^2)^2 / (2\hbar^2)$. (Give five significant figures in both answers.) The difference in your answers is small enough that we are usually

justified in ignoring the proton's motion. Nevertheless, the difference can be detected, and the result (5.43) is found to be correct.

SECTION 5.9 (X-ray Spectra)

- 5.22** • Use Equation (5.37) to predict the slope of a graph of \sqrt{f} against Z for the K_α frequencies. Do the data in Fig. 5.7 bear out your prediction?

- 5.23** • The K_α line from a certain element is found to have wavelength 0.475 nm. Use Equation (5.37), with $\delta \approx 1$, to determine what the element is.

- 5.24** • The K series of X-rays consists of photons emitted when an electron drops from the n th Bohr orbit to the first ($n \rightarrow 1$). (a) Use (5.33) to derive an expression for the wavelengths of the K series. [This will be approximate, since (5.33) ignores effects of screening.] (b) Find the wavelengths of the K_α , K_β , and K_γ , lines ($n = 2, 3, 4$) of uranium. (For the atomic numbers of uranium and other elements, see the periodic table inside the back cover or the alphabetical lists in Appendix C.)

- 5.25** • What is the approximate radius of the $n = 1$ orbit of the innermost electron in the lead atom? Compare with the radius of the lead nucleus, $R \approx 7 \times 10^{-15}$ m.

- 5.26** • Suppose that a negative muon (see Problem 5.13) penetrates the electrons of a silver atom and is captured in the first Bohr orbit around the nucleus. (a) What is the radius (in fm) of the muon's orbit? Is it a good approximation to ignore the atomic electrons when considering the muon? (The muon's orbital radius is very close to the nuclear radius. For this reason, the details of the muon's orbit are sensitive to the charge distribution of the nucleus, and the study of muonic atoms is a useful probe of nuclear properties.) (b) What are the energy and wavelength of a photon emitted when a muon drops from the $n = 2$ to the $n = 1$ orbit?

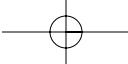
SECTION 5.10 (Other Evidence for Atomic Energy Levels)

- 5.27** ••• When an electron with initial kinetic energy K_0 scatters elastically from a stationary atom, there is no loss of total kinetic energy. Nevertheless, the electron loses a little kinetic energy to the recoil of the atom.

- (a) Use conservation of momentum and kinetic energy to prove that the maximum kinetic energy of the recoiling atom is approximately $(4m/M)K_0$, where m and M are the masses of the electron and atom.

[HINT: The maximum recoil energy is in a head-on collision. Remember that $m \ll M$, and use nonrelativistic mechanics.]

- (b) If a 3-eV electron collides elastically with a mercury atom, what is its maximum possible loss of kinetic energy? (Your answer should convince you that it is a good approximation to say that the electrons in the Franck–Hertz experiment lose no kinetic energy in elastic collisions with atoms.)



Chapter 6

Matter Waves

- 6.1 Introduction
- 6.2 De Broglie's Hypothesis
- 6.3 Experimental Verification
- 6.4 The Quantum Wave Function
- 6.5 Which Slit Does the Electron Go Through?
- 6.6 Sinusoidal Waves
- 6.7 Wave Packets and Fourier Analysis
- 6.8 The Uncertainty Relation for Position and Momentum
- 6.9 The Uncertainty Relation for Time and Energy
- 6.10 Velocity of a Wave Packet *

Problems for Chapter 6

* This section can be omitted without serious loss of continuity.

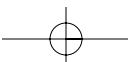
6.1 Introduction

Bohr published his model of the hydrogen atom in 1913. Although it was studied intensely during the next 10 years, there was little progress toward a complete theory that could explain the model's success with hydrogen or could give a satisfactory account of multielectron atoms. Then, in 1923, a French doctoral student named Louis de Broglie proposed an idea that gave a new understanding of the Bohr model and proved to be the essential step in the development of modern quantum mechanics.

Appealing to the hope that nature is symmetric, de Broglie reasoned that if light has both wavelike and particle-like properties, material objects such as electrons might also exhibit this dual character. At that time there was no known evidence for wavelike properties of any material particles. Nevertheless, de Broglie showed that if electrons were assumed to behave like waves, Bohr's stationary orbits could be explained as standing waves inside the hydrogen atom. These proposed waves, whose exact nature did not become clear for another two or three years, came to be called "matter waves".

De Broglie's idea that particle-wave duality applied to electrons as well as photons appealed to many physicists. It was taken up by the Austrian physicist Erwin Schrödinger, whose four papers published in 1926 mark the birth of modern *wave mechanics*, or *quantum mechanics* as we usually say today.* The

*At about the same time, Heisenberg developed an independent form of quantum mechanics, which was later shown to be equivalent to Schrödinger's theory. We have chosen to describe Schrödinger's approach because it is more easily understood.



Section 6.2 • De Broglie's Hypothesis 169

next year, 1927, saw direct experimental verification of de Broglie's matter waves, with the observation of interference patterns (the hallmark of any wave phenomenon) produced by electrons.

In this chapter we discuss the properties of de Broglie's matter waves and their experimental verification. We write down de Broglie's relations for the frequency and wavelength of matter waves, and we discuss their physical interpretation — suggested by the German physicist Max Born in 1926 — as waves whose intensity at any point gives the *probability* of finding the particle at that point. In Chapter 7 we introduce the Schrödinger equation, which is the equation of motion for matter waves and plays the same role in quantum mechanics as Newton's second law plays in classical mechanics.

6.2 De Broglie's Hypothesis

We saw in Chapter 4 that photons display the properties of both waves and particles and that the two kinds of properties are related by the equations (4.27)

$$E = hf \quad \text{and} \quad p = \frac{h}{\lambda} \quad (6.1)$$

De Broglie proposed that material particles, such as electrons, should show a similar particle-wave duality. He did not know the precise nature of the proposed matter waves, but he argued that they should satisfy the same two relations (6.1) as apply to light waves.* For this reason, the relations (6.1), as applied to matter waves, are often called the **de Broglie relations**.

At the time there was no known experimental evidence for de Broglie's proposed matter waves. However, one can see (at least in retrospect) that electron waves provide a plausible explanation for the quantization of atomic energy levels. If we accept the de Broglie relations (6.1), quantization of the electron's energy E is equivalent to quantization of the frequency f of the electron wave. It is a familiar fact from classical physics that waves which are confined in some region (sound waves in an organ pipe, waves on a stretched string) can vibrate only at certain discrete, quantized frequencies. This suggests that quantization of atomic energy levels might be explained as the quantization of the frequency of the electron waves confined inside the atom.

In fact, de Broglie was able to argue that if electrons were some kind of wave satisfying the relations (6.1), the angular momentum of an electron in a hydrogen atom would be quantized in multiples of \hbar exactly as required by the Bohr model. He pictured the electron wave as somehow vibrating around the Bohr orbit, as shown in Fig. 6.1. It is evident that a wave can fit onto a circular path of radius r as shown only if the circumference can accommodate an integral number of wavelengths; that is,

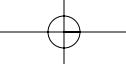
$$2\pi r = n\lambda \quad n = 1, 2, 3, \dots \quad (6.2)$$

*This suggestion was not as obvious as we may have made it seem. In the case of light waves, whose speed is c , the relations (6.1) can be rewritten in various equivalent ways. For example, since $f = c/\lambda$ (for light), the relation $E = hf$ can be rewritten as $E = hc/\lambda$. Since matter waves do not have speed c , this second form is *incorrect* for matter waves. De Broglie was led to the correct relations (6.1) mainly by considerations of relativistic invariance.

Louis De Broglie
(1892–1987, French)



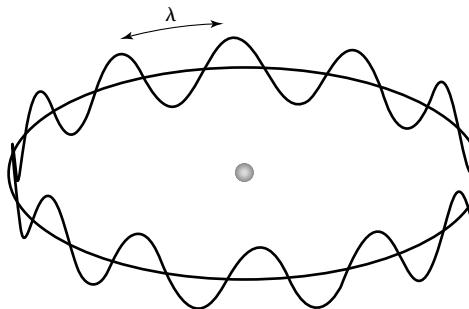
The son of a noble French family, one of whose ancestors had been guillotined in the French Revolution, de Broglie graduated from the Sorbonne in history. He became interested in physics during World War I, and his idea that all particles have an associated wave was in his PhD thesis in 1924. This proposal, which earned him the 1929 Nobel Prize in physics, was an essential link between the Bohr model and modern quantum mechanics.



170 Chapter 6 • Matter Waves

FIGURE 6.1

If an electron wave — whatever it may be — is pictured as circling around the atomic nucleus, its wavelength λ must fit an integer number of times into the circumference.



Now, according to (6.1) $\lambda = h/p$. Thus (6.2) implies that $2\pi r = nh/p$ or

$$rp = \frac{nh}{2\pi} \quad (6.3)$$

But for a circular orbit (which we are considering), rp is just the angular momentum L , and we conclude that

$$L = \frac{nh}{2\pi} = n\hbar \quad n = 1, 2, 3, \dots$$

which is just the Bohr quantization condition.

We should emphasize that this explanation of the quantization of angular momentum is no longer considered completely satisfactory. For instance, the wave shown in Fig. 6.1 is some kind of one-dimensional wave, constrained to move around a circular path with a definite radius r , whereas modern quantum mechanics envisions a three-dimensional wave that is spread throughout the whole atom (more like an acoustic wave reverberating inside a chamber). Nevertheless, the central idea — that quantization of angular momentum was explained by the notion of electron waves satisfying the de Broglie relations — was absolutely correct.

6.3 Experimental Verification

If electrons and other material particles have wave properties, as de Broglie suggested, the question naturally arose why these wave properties had never been observed. The answer lies in the extremely short wavelength of most matter waves. You will recall that the wave nature of light was only firmly established in the nineteenth century by the observation of interference patterns in Young's double-slit experiment and other similar experiments. What made those experiments difficult was that the wavelength of light ($\lambda \approx 400$ to 700 nm) is very small by everyday macroscopic standards. To obtain an interference pattern that is easily observed, one must make the distance between the slits small enough that the path lengths from the two slits to points on the screen differ by about a wavelength, and this is not easy to do. In the case of de Broglie's matter waves, the wavelengths are usually much shorter than those of visible light, as we see in Example 6.1 below. Thus, the wave nature of material particles was even more difficult to observe than was that of light.

Example 6.1

Use the de Broglie relations (6.1) to find the wavelengths of electrons with kinetic energies $K = 10, 100, 1000$, and $10,000$ eV. Compare these with the wavelengths of visible light and X-rays. Would a heavier particle, such as a proton, with the same energies have longer or shorter wavelengths?

The wavelength is given by the relation $\lambda = h/p$. Thus, our first task is to express the momentum p in terms of the kinetic energy K .^{*} Since all energies concerned are small compared to the electron's rest energy, $mc^2 \approx 0.5$ Mev, we can use the nonrelativistic expression

$$K = \frac{1}{2}mv^2 = \frac{1}{2}\frac{p^2}{m}$$

which implies that

$$p = \sqrt{2mK}$$

Therefore,

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mK}} \quad (6.4)$$

For an electron with $K = 10$ eV and $mc^2 = 0.51$ MeV, this gives

$$\lambda = \frac{hc}{\sqrt{2mc^2K}} = \frac{1240 \text{ eV} \cdot \text{nm}}{\sqrt{2 \times (0.51 \times 10^6 \text{ eV}) \times (10 \text{ eV})}} = 0.39 \text{ nm}$$

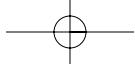
Since λ is inversely proportional to \sqrt{K} , we can immediately write down λ for all four energies as follows:

K (eV):	10	100	1000	10,000
λ (nm):	0.39	0.12	0.039	0.012

All of these wavelengths are very much shorter than those of visible light (400 to 700 nm); they span the range of X-ray wavelengths. From (6.4) it is clear that for a given energy K , the wavelength λ is inversely proportional to \sqrt{m} . Thus, the heavier the particle, the shorter will be the wavelength, and for a given energy, the best chance of observing matter waves is with electrons.

Since the wavelength of electrons in the range 100 to 1000 eV is comparable to that of the X-rays used in X-ray diffraction from crystals, de Broglie suggested that it might be possible to observe diffraction of electron waves by crystals. Unfortunately, a beam of electrons with kinetic energies of only a few

*An alternative might seem to be to find the frequency from the relation $E = hf$ and then the wavelength from $\lambda = v_{\text{wave}}/f$, where v_{wave} denotes the speed of the matter wave. Unfortunately, we do not know v_{wave} , which cannot be assumed to be the same as the speed of the particle. See Section 6.10.



172 Chapter 6 • Matter Waves



FIGURE 6.2

The electron tube in which Davisson and Germer observed the diffraction of electron waves. Notice the graduated turntable for rotating the target crystal near the center of the tube.

hundred eV requires an extremely good vacuum to avoid severe scattering of the electrons by the remaining gas in the electron tube. For this reason, several attempts to observe electron diffraction failed, and it was not until 1927 that the American physicists Clinton Davisson and Lester Germer published conclusive evidence for the diffraction of electron waves. Working in the Western Electric Laboratories in New York (later to become Bell Labs), they directed a beam of 54-eV electrons at a crystal of nickel. They found numerous maxima and minima in the scattered intensity at angles that were consistent with the diffraction of waves with wavelength, $\lambda = h/p$. A photograph of Davisson and Germer's electron tube is shown in Fig. 6.2. Their experiment left no doubt of the existence of electron waves and of the correctness of the de Broglie relation $\lambda = h/p$.

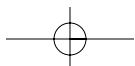
In the same year, G. P. Thomson (son of J. J. Thomson, discoverer of the electron) demonstrated diffraction of electrons transmitted through thin metal foils. The younger Thomson and Davisson shared the 1937 Nobel Prize for their discovery of matter waves. Within a few years, diffraction of several other particles — hydrogen atoms, helium atoms, and later neutrons — had been observed, and it was reasonably clear that de Broglie's ideas applied to all material particles. Figure 6.3 shows diffraction patterns made by X-rays, electrons, and neutrons, transmitted through polycrystalline metals. The first picture is the same picture shown in Figure 4.10 as evidence for the wave nature of X-rays. The similarity of the three patterns is unmistakable evidence that electrons and neutrons are also wave phenomena.

Since one of the first and best known demonstrations that light is a wave was Young's two-slit experiment, it is interesting that this same experiment has



FIGURE 6.3

Diffraction rings produced by diffraction of waves in polycrystalline metal samples with (a) X-rays, (b) electrons, (c) neutrons.



**FIGURE 6.4**

Two-slit interference patterns produced by light and electrons.

now been carried out with electron waves. This requires extremely narrow slits, very close together, and the pattern has to be enlarged many times to be discernible. Figure 6.4 shows photographs of two-slit patterns made with beams of light and electrons. Here, again, the similarity is unmistakable.

Today we take for granted that all material particles have wave properties with wavelength and frequency given by the de Broglie relations, and these wave properties have found many applications. The short wavelength of electron waves is exploited in the electron microscope. An ordinary microscope uses light (focused by glass lenses) and cannot resolve objects much smaller than 10^{-6} m. An electron microscope uses electrons, focused by magnetic fields; because the electrons have wavelengths thousands of times smaller than light, the electron microscope can resolve objects down to about 10^{-10} m*.

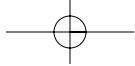
The diffraction phenomena that originally established the wave properties of matter are now used as probes of the structure of solids. As we saw, electrons with suitable wavelength have energies of 100 eV or so — much lower than the energy of X-rays with the same wavelength. These low-energy electrons do not penetrate as deeply into matter as do the X-rays, and low-energy electron diffraction (or LEED) is therefore used to study surface properties of solids. Since these surface properties are important in electronic and catalytic devices, LEED has become a widely used research tool. See Chapter 14 for a little more on surface science.

It has also proved possible to work with neutrons of very low energy — a few hundredths of an eV. Even though the neutron is so much heavier than the electron, its wavelength at these low energies is comparable with that of X-rays, and neutron diffraction is another important probe of the structure of solids. One advantage of using neutrons is that they are scattered appreciably by hydrogen, since they are subject to the nuclear force of the proton; X-rays and electrons are scattered only weakly by hydrogen since they interact mainly with electric charge (of which the hydrogen atom contains relatively little — one electron and one proton). Thus, neutron diffraction is often the most effective way to study crystals containing hydrogen.

6.4 The Quantum Wave Function

For a complete description of any wave, we must discuss its **wave function**. This is the mathematical function that specifies the wave disturbance at each point of space and time. For waves on a taut string, aligned along the x axis, the wave

*The wavelength of high-energy electrons in electron microscopes (typical energies 50–500 keV) is much smaller than 10^{-10} m, as shown in Example 6.1. The resolution of electron microscopes is limited, not by the electron wavelength, but by unavoidable aberrations in the magnetic lenses which focus the electrons and form the image.



174 Chapter 6 • Matter Waves

function $y(x, t)$ gives the transverse displacement of the string, for all positions x and times t . For sound waves, the wave function is the pressure change p resulting from the wave. Since sound waves usually travel outward in all directions, p depends on all three spatial coordinates and time,

$$p = p(x, y, z, t) = p(\mathbf{r}, t)$$

For light waves, the wave function is the electric field strength $\mathcal{E}(\mathbf{r}, t)$. De Broglie had no clear idea what his matter waves really were; in other words, he did not know the nature of their wave function. The interpretation of the matter wave function that is generally accepted today was proposed by the German physicist Max Born in 1926. Born's ideas were taken up and extended by Bohr and his associates in Copenhagen and, for this reason, are often called the Copenhagen interpretation of quantum mechanics. The distinctive feature of Born's proposal is that the matter wave function specifies only *probabilities* — rather than specific values — of a particle's properties, as we now describe.

To understand Born's proposal, it is helpful to consider, as Born did, the connection between the electromagnetic wave function $\mathcal{E}(\mathbf{r}, t)$ and the photon. It was known in classical electromagnetism that the electric field strength determines the energy carried by an electromagnetic wave. Specifically, the energy E in any small volume dV at a point \mathbf{r} (and time t) is

$$E(\text{in volume } dV \text{ at } \mathbf{r}) = \epsilon_0 [\mathcal{E}(\mathbf{r}, t)]^2 dV \quad (6.5)$$

where ϵ_0 is the constant called the permittivity of the vacuum. To avoid having to worry about constants like ϵ_0 we will rewrite (6.5) as a proportion.

$$E(\text{in volume } dV \text{ at } \mathbf{r}) \propto [\mathcal{E}(\mathbf{r}, t)]^2 dV \quad (6.6)$$

From a quantum point of view, we know that the energy of an electromagnetic wave is carried by discrete photons. If, for simplicity, we consider a wave with a single fixed frequency f , each photon has energy hf . Therefore, we can divide (6.6) by hf to give

$$\text{(number of photons in } dV \text{ at } \mathbf{r}) = \frac{E(\text{in volume } dV \text{ at } \mathbf{r})}{hf} \propto [\mathcal{E}(\mathbf{r}, t)]^2 dV \quad (6.7)$$

(Note that the proportionality sign lets us omit the constants h and f .) Since the square of any wave function is often called the *intensity*, we can paraphrase (6.7) to say that the number of photons in a small volume dV is proportional to the intensity $[\mathcal{E}(\mathbf{r}, t)]^2$ of the light.

The result (6.7) cannot be exactly true as written. If we were to choose a small enough volume dV , then (6.7) would predict a fractional number of photons, and this is impossible. A correct statement is that (6.7) gives the *probable number* of photons in the volume dV

$$\text{(probable number of photons in } dV \text{ at } \mathbf{r}) \propto [\mathcal{E}(\mathbf{r}, t)]^2 dV \quad (6.8)$$

To illustrate what this means, suppose that we shine a steady beam of light across a room. We select some definite small volume dV and imagine somehow counting the number of photons in dV (at any instant t). The “probable number” given by (6.8) is the *average result* expected if we repeat this same counting experiment many times. Suppose, for example, the number predicted

Max Born
(1882–1970, German-British)



Born is best known for his work on the mathematical structure of quantum mechanics and the interpretation of the wave function. After getting his doctorate, he worked with J.J. Thomson at Cambridge and lectured in Chicago for Michelson. He subsequently became professor at Berlin and then Göttingen. When Hitler came to power, Born left Germany and became a professor at Edinburgh University. He won the Nobel Prize in physics in 1954 for his contributions to quantum theory.

Section 6.4 • The Quantum Wave Function **175**

by (6.8) for a certain small volume dV is 1.5. Of course, we cannot find 1.5 photons in dV in any one observation. On the contrary, each observation will yield some whole number, and we might get 1 photon, then 2, then 0, and so on. After many repeated observations, our *average* result will be 1.5.

Born proposed that a relation similar to (6.8) should apply to electron waves (and any other matter waves); that is, there must be an electron wave function, usually denoted by the Greek capital letter psi*, $\Psi(\mathbf{r}, t)$, whose square gives the probable number of electrons in a small volume dV .

$$(\text{probable number of electrons in } dV \text{ at } \mathbf{r}) \propto [\Psi(\mathbf{r}, t)]^2 dV \quad (6.9)$$

A small but important complication with matter waves is that the wave function Ψ has two parts, which can be conveniently expressed by use of complex numbers: The value of Ψ (at any point) can be written as a complex number, with real and imaginary parts:

$$\Psi = \Psi_{\text{real}} + i\Psi_{\text{imag}} \quad (6.10)$$

where i is the imaginary number, $i = \sqrt{-1}$. We will see later, in Section 7.3, that this unexpected feature of matter waves is central to the modern explanation of Bohr's stationary orbits. With Ψ complex, (6.9) is modified to read

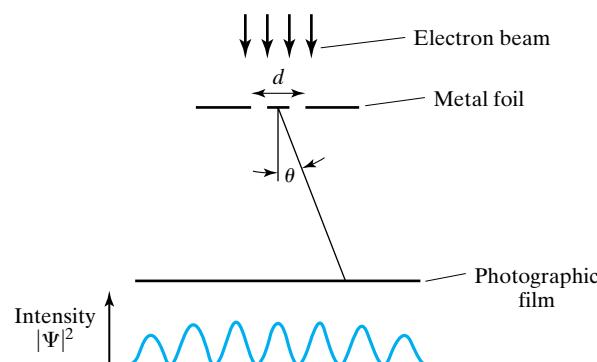
$$(\text{probable number of electrons in } dV \text{ at } \mathbf{r}) \propto |\Psi(\mathbf{r}, t)|^2 dV \quad (6.11)$$

where $|\Psi|$ denotes the absolute value of the complex number (6.10),

$$|\Psi| = \sqrt{\Psi_{\text{real}}^2 + \Psi_{\text{imag}}^2} \quad (6.12)$$

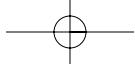
We will refer to $|\Psi|^2$ as the intensity of the wave function. Since $|\Psi|$ is a real number, the intensity is real and positive, and (6.11) guarantees that the probable number of electrons is positive, as it has to be.

To better understand the significance of the matter (or quantum) wave function Ψ , let us consider in detail an interference experiment such as the two-slit experiment. As we have seen, this can be done with electrons or photons, but we will consider the case of electrons. A beam of electrons is directed at a metal foil with two narrow slits, as shown in Fig. 6.5. The electron waves that pass through the slits can interfere constructively and destructively on the far side. The resulting interference pattern can be recorded by placing a photographic film, as shown, since each electron striking the film will leave a

**FIGURE 6.5**

Schematic diagram of the two-slit experiment. The graph shows the intensity as a function of position along the film.

*Pronounced "sigh."



176 Chapter 6 • Matter Waves

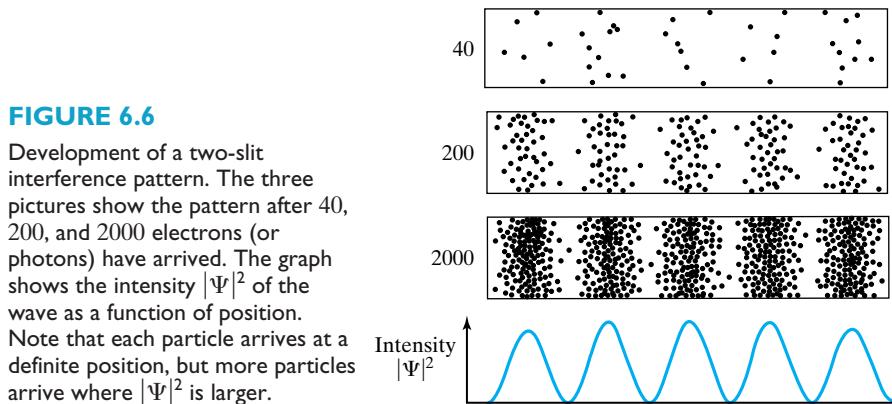


FIGURE 6.6

Development of a two-slit interference pattern. The three pictures show the pattern after 40, 200, and 2000 electrons (or photons) have arrived. The graph shows the intensity $|\Psi|^2$ of the wave as a function of position. Note that each particle arrives at a definite position, but more particles arrive where $|\Psi|^2$ is larger.

small dark spot. The directions in which constructive interference occurs are given by the well-known condition

$$d \sin \theta = n\lambda, \quad n = 0, \pm 1, \pm 2, \dots$$

where d is the separation of the slits and λ the wavelength of the wave. The resulting intensity, as a function of position along the screen, is plotted at the bottom of Fig. 6.5.

According to (6.11), we expect many electrons to arrive at those points where the intensity $|\Psi|^2$ is large and very few to be seen where $|\Psi|^2$ is small. This is exactly what is found, as illustrated in Fig. 6.6, which shows a typical two-slit pattern after 40, 200, and 2000 electrons have arrived. After just 40 electrons, no obvious pattern stands out, but it is clear that each electron, as represented by a black dot on the film, arrives at a definite position. After 200 electrons, the characteristic two-slit pattern is becoming discernible; after 2000, the pattern is obvious and we see clearly that most of the electrons arrive near those points where $|\Psi|^2$ is maximum.

If we use a weak beam of electrons, the patterns of Fig. 6.6 will develop slowly. But even if the beam is so weak that only one electron is in the apparatus at a time, the interference pattern still appears. This implies that there is a quantum wave associated with *each individual* electron. This wave passes through the two slits, and the resulting interference determines where the electron is most likely to be found. Each electron arrives at a definite spot, but after many electrons have arrived, we can identify the regions where $|\Psi|^2$ is greatest as the regions where more electrons have arrived.

These considerations let us sharpen our interpretation of the quantum wave function: Associated with each individual quantum particle there is a wave function $\Psi(\mathbf{r}, t)$, whose intensity at any position \mathbf{r} determines the probability P of finding the particle at \mathbf{r} (at time t):

$$P(\text{finding particle in } dV \text{ at } \mathbf{r}) \propto |\Psi(\mathbf{r}, t)|^2 dV \quad (6.13)$$

Since P is proportional to $|\Psi|^2 dV$, we can choose the scale of Ψ so that P equals $|\Psi|^2 dV$, that is, we choose the units of Ψ so that the constant of proportionality in (6.13) has the value 1. With this choice we can rewrite (6.13) as follows: If $\Psi(\mathbf{r}, t)$ is the wave function associated with a quantum particle, then at any time t

$$|\Psi(\mathbf{r}, t)|^2 dV = P(\text{finding particle in } dV \text{ at } \mathbf{r}) \quad (6.14)$$

This is Born's interpretation of the quantum wave function $\Psi(\mathbf{r}, t)$: It is a wave whose intensity gives the probability of finding the particle at \mathbf{r} . Another way to express (6.14) is to divide both sides by the volume dV , in which case the right-hand side becomes the **probability density** (that is, the probability per unit volume). We can then say that at any time t

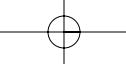
$$|\Psi(\mathbf{r}, t)|^2 = \text{probability density for finding particle at } \mathbf{r} \quad (6.15)$$

It is important to understand what (6.14) and (6.15) assert. Suppose that we can somehow arrange an experiment with a succession of quantum particles, all with the same wave function Ψ . If we measure *one* particle's position (by letting it run into a photographic film, for example), we will find some definite result (in the form of a single dark spot on the film). However, after measuring several particles, all with exactly the same wave function, we will generally get several different answers. What the quantum wave function Ψ tells us is the frequency of occurrence of the various possible positions. The particles will be found most often at those points where $|\Psi|^2$ is largest and least often where $|\Psi|^2$ is smallest, in accordance with (6.14) or (6.15).

Even when we know a particle's wave function Ψ exactly, we cannot specify a unique position of the particle; rather, we can give only the respective probabilities that the particle will be found at the various possible positions. For this reason, modern quantum mechanics is often described as a probabilistic theory. This is perhaps the most profound difference between classical and quantum mechanics. In classical mechanics, every experiment has an outcome that can, in principle, be predicted unambiguously; in quantum mechanics the same experiment repeated under the same conditions can produce different outcomes. Nonetheless, the probabilities of the various results *can* be predicted, and it is these probabilities that are measured in an experiment such as the two-slit experiment.

Equation (6.14) gives the answer to the question: What is the wave function Ψ of a quantum particle? According to (6.14), Ψ is a function whose intensity $|\Psi|^2$ gives the probability of finding the particle at any particular position. The quantum wave function is much harder to visualize than classical wave functions, such as the displacement $y(x, t)$ produced by a wave on a string. Nevertheless, the probabilistic nature of (6.14) is an essential characteristic of the quantum wave function, and is one of the properties on which modern quantum mechanics is built.

The probabilistic character of quantum mechanics appears in almost all quantum processes. Consider, for example, the scattering of X-rays by electrons in the Compton effect. Quantum mechanics allows one to calculate the *probability* that an X-ray photon will be scattered by an electron. Thus, if we send in many identical photons, quantum mechanics predicts unambiguously the fraction of photons scattered, but it *cannot* predict the fate of any single photon. This can be likened to the situation in an everyday toss of a coin: Probability theory tells us that if we toss many identical coins, 50% will land "heads," but the theory *cannot* tell us the fate of any individual coin. This analogy is not perfect, however. The probabilistic character of the coin toss arises from our ignorance of the details of the toss. If we knew the precise conditions of the coin's launch, then — in principle — we *could* predict the fate of a single coin. By contrast, the probabilistic character of quantum mechanics is an inherent feature of the theory. Even though we know *all* there is to know about the electron, its position cannot be predicted. Its precise position does not exist unless and until a position measurement is made.



178 Chapter 6 • Matter Waves

6.5 Which Slit Does the Electron Go Through?

When an electron passes through the two-slit apparatus, its wave must pass through both slits in order to produce interference. Thus if we observe interference, we can be sure that the wave certainly passed through *both* slits. Nevertheless, it is natural to ask which slit the *electron itself* went through. Reasonable as this question seems, it is a question without an answer. Since the wave had the same intensity in each slit, there is an equal probability that we would have found it in either slit if we had placed suitable detectors in the slits; but if we conduct the experiment in the usual way, without any such detectors, it is impossible to say which slit the electron actually went through. Furthermore, if we do use detectors to monitor which slit the electron uses, the detectors will disturb the electron waves as they pass through, and the two-slit pattern will disappear.

A general proof of this last statement is beyond the scope of this book, and we will have to be content with illustrating it by means of a simple thought experiment. Let us suppose that we decide to find out which slit each electron passes through by shining a narrow beam of light through one slit, as in Fig. 6.7. Each time an electron passes through that slit, it will scatter a few photons and can be detected by a brief reduction in the light transmitted through the slit. In this way, we can determine (in principle, at least) which slit each electron passes through. Unfortunately, we acquire this knowledge at the cost of having the electron collide with at least one photon; and this collision disturbs the electron's motion enough to destroy completely the two-slit interference pattern, as we now argue.

In order to observe a two-slit pattern it is convenient to use electrons with wavelength of the same order as the slit separation d . We can write this requirement as*

$$\lambda_{\text{el}} \approx d$$

which implies that

$$p_{\text{el}} = \frac{h}{\lambda_{\text{el}}} \approx \frac{h}{d} \quad (6.16)$$

To distinguish between the two slits, we must make our beam of light appreciably narrower than d , and it is a well-known result that this is possible only if the wavelength λ_γ of the light is of order d or less:

$$\lambda_\gamma \lesssim d$$

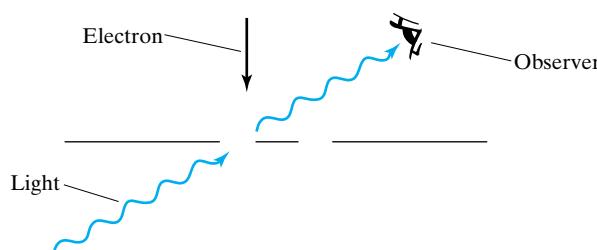


FIGURE 6.7

To find out which slit the electron passes through, we shine a narrow beam of light through one of the slits.

*For the sake of brevity, we consider only the case that λ_{el} is of order d . If λ_{el} is much less than d (which is possible), the argument becomes slightly more complicated. However, the conclusion is the same.

This requires that

$$p_\gamma = \frac{h}{\lambda_\gamma} \gtrsim \frac{h}{d} \quad (6.17)$$

Comparing (6.16) and (6.17), we see that the photons used to detect the electrons must have momentum at least of the same order as that of the electrons themselves. Under these conditions, a single collision with a photon is enough to change, completely and randomly, the electron's momentum. After the passage of many electrons, each of which has been disturbed in this random fashion, there will be no interference pattern. By using a beam of light to see which slit each electron traversed, we have destroyed the two-slit interference pattern.

This conclusion is independent of the details of the experiment (see Problem 6.17) and illustrates an important difference between classical and quantum mechanics. According to classical mechanics, the electron follows a definite path and hence must pass through one slit or the other in a predictable manner. In quantum mechanics precise statements *can* be made about the probabilities of finding the electron at various positions, but one unique position *cannot* be predicted. Indeed, when one uses a detector to measure the electron's position, subsequent predictions of its position are radically altered. The need to analyze carefully the process of measurement and the disturbances it produces was first pointed out by Heisenberg, who was led by these considerations to his famous uncertainty principle, which we describe in Sections 6.8 and 6.9.

6.6 Sinusoidal Waves

The de Broglie relations, $E = hf$ and $p = h/\lambda$, imply that if a particle has definite values for its energy and momentum, its wave function has corresponding definite values for its frequency and wavelength. A wave with definite frequency and wavelength is called a *sinusoidal*, or *harmonic*, *wave*. In this section we review briefly the properties of these waves. As we will argue later, an *exactly* sinusoidal wave is an idealization that never really occurs in practice. Nevertheless, many real waves are well approximated by sinusoidal waves, and as we will describe in Section 6.7, *any* wave can be built up as a sum of sinusoidal waves. Thus, it is important to be familiar with their properties.

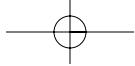
To simplify our discussion, we will consider the simplest of classical waves, namely waves on a string. However, most of the ideas of this and the next section apply with only small changes to any wave, classical or quantum.

Let us consider first a sinusoidal wave traveling to the right on a taut string. The wave function for such a wave has the form

$$y(x, t) = A \sin 2\pi \left(\frac{x}{\lambda} - \frac{t}{T} \right) \quad (6.18)$$

The term "sinusoidal" is used to describe a wave given by either a sine function (6.18) or a cosine function since both have the same general characteristics. In (6.18), $y(x, t)$ is the transverse displacement of the string at position x along the string and time t . The constants A , λ , and T are called the amplitude, wavelength, and period.

A function like (6.18) that depends on two variables x and t is hard to visualize. One way to describe it is by a sequence of snapshots taken at equally



180 Chapter 6 • Matter Waves

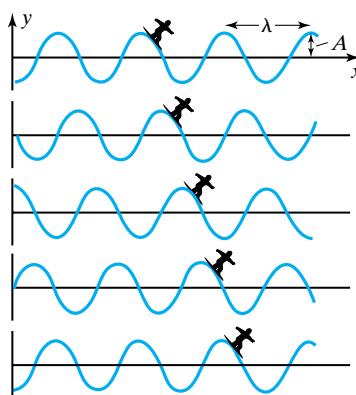


FIGURE 6.8

Five successive snapshots of the sinusoidal wave (6.18). Any definite point on the wave, like the crest on which the surfer is riding, moves steadily to the right. At any fixed position x , the string bobs up and down sinusoidally in time.

spaced times as in Fig. 6.8. Each picture shows the displacement y as a function of x for one definite time t , and for the wave (6.18) each snapshot has the shape of a sine function. As time goes by, the whole wave moves steadily to the right, as can be seen by focusing on the surfer shown riding on one wave crest. If, instead, we focus on one particular position x , the string oscillates up and down as a sinusoidal function of time. Thus, a graph of y against t for any fixed x would have the same general shape as any one of the pictures in Fig. 6.8.

The wave (6.18), with a minus sign in its argument, travels to the right. If we replace the minus sign with a plus, we get a wave,

$$y(x, t) = A \sin 2\pi \left(\frac{x}{\lambda} + \frac{t}{T} \right)$$

that travels to the left. (See Problem 6.26.)

The significance of the constants A and λ is clear in Fig. 6.8. The amplitude A is the maximum displacement of the string from its mean position. The wavelength λ is the distance one must move (at one fixed time t) before the wave repeats itself. The period T is the time one must wait (at one fixed point x) for the wave to repeat itself. The frequency f is the number of oscillations in unit time at one fixed point and is given by $f = 1/T$. The frequency is usually measured in s^{-1} , or hertz (1 hertz $\equiv 1$ Hz $\equiv 1$ s $^{-1}$).

It is often convenient to rewrite the wave function (6.18) as

$$y(x, t) = A \sin(kx - \omega t) \quad (6.19)$$

where k is called the **wave number** (measured in rad/m) and ω the **angular frequency** (in rad/s). Comparing (6.18) and (6.19), we can express k and ω in terms of λ and T (or f). The various parameters characterizing a wave can be summarized as follows: The spatial parameters are

wavelength, λ

$$\text{wave number, } k = \frac{2\pi}{\lambda} \quad (6.20)$$

and the parameters related to time

period, T

$$\text{frequency, } f = \frac{1}{T} \quad (6.21)$$

$$\text{angular frequency, } \omega = 2\pi f = \frac{2\pi}{T}$$

The speed with which any particular crest (or any other definite point on the wave) moves is the wave speed

$$v = \frac{\lambda}{T} = \lambda f = \frac{\omega}{k} \quad (6.22)$$

since the wave moves a distance λ in time T . In quantum mechanics the parameters k and ω are used frequently, and it is a good idea to be familiar with the de Broglie relations in terms of them:

$$E = hf = \hbar\omega \quad (6.23)$$

and

$$p = \frac{h}{\lambda} = \hbar k \quad (6.24)$$

6.7 Wave Packets and Fourier Analysis

A sinusoidal wave, with definite frequency and wavelength, is a mathematical idealization that never occurs in practice. For example, we speak of a pure musical tone as a harmonic wave with one precise frequency, but a careful analysis shows that any real musical note is a mixture of many different frequencies. Similarly, we speak of monochromatic light as light with a single well-defined frequency, but even light from the most monochromatic laser is found on close inspection to have a small spread of frequencies.

It is easy to see why an exact sine wave cannot occur. A pure sine wave like (6.19) is perfectly periodic, repeating itself endlessly in time and in space. A real wave may repeat itself for a long time and over a large distance, but certainly not *forever*. This is especially clear in the case of matter waves. A wave function that was a pure sine wave would extend indefinitely and would represent a particle that was equally likely to be found *anywhere*. This is obviously absurd. Imagine, for example, an electron in a TV tube. At some time t_0 it is ejected from the cathode, and at some later time t_1 it hits the screen, creating a small flash of light. At any time between t_0 and t_1 , the electron — being a wave — cannot have a precisely defined position. Nevertheless, we know from experience that the electron is definitely within some region that is tiny compared to the size of the TV tube. When t is close to t_0 , this region is near the cathode; and as t advances, the region moves toward the screen. Thus, at any instant during its flight, the electron's wave function might look something like Fig. 6.9. The important points about this wave function are that (1) there is a region, labeled $\pm\Delta x$, where the wave differs from zero and the electron may be found, and that (2) outside this region the wave is zero (or very small) and the electron will not be found (or is very unlikely to be found). We refer to this kind of wave function that is *localized* within some region as a **wave packet** or **wave pulse**.

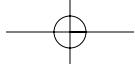
Our aim in this section is to describe the mathematical properties of localized wave packets like that shown in Fig. 6.9. We will find that they can be built up as superpositions of many different sinusoidal waves. We begin our discussion with a crash course in **Fourier analysis**, the study of the composition of functions as sums of sinusoidal waves, named after the French mathematician Joseph Fourier. The central result of Fourier analysis is that any function* can be written as a sum of sines and cosines. We consider two broad categories of functions: (1) periodic functions, that is, those that are cyclical and repeat at regular intervals and (2) nonperiodic, nonrepeating functions, such as the wave packet of Figure 6.9.

Joseph Fourier
(1768–1830, French)



As a young man, Fourier couldn't decide whether to devote his life to mathematics, the priesthood, or politics. He finally became a politically active mathematician. Acting as scientific advisor, Fourier accompanied Napoleon's army to Egypt in 1798, where he conducted archeological expeditions and established educational institutions. He was a prolific and well-respected mathematician; his most famous work was *On the Propagation of Heat in Solid Bodies*, written in 1807. In this book he developed the technique now called Fourier analysis and applied it to problems of heat flow. The work was controversial, and many eminent mathematicians refused to believe that functions with discontinuities could be written as sums of smoothly varying sines and cosines.

*That is, any function that would interest a physicist. There are “pathological” functions, for example, those with an infinitely dense number of discontinuities, which cannot be written in terms of sines and cosines.



182 Chapter 6 • Matter Waves

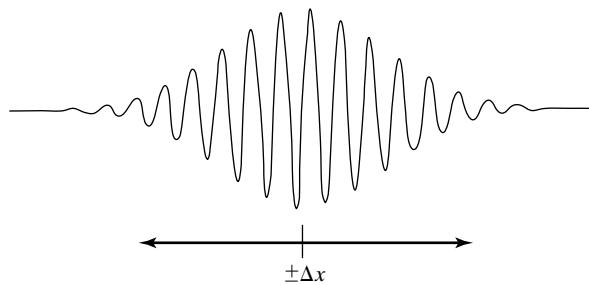


FIGURE 6.9

A localized wave packet, which is nonzero in an interval $\pm\Delta x$ and zero elsewhere. A particle with this wave function would have a vanishingly small probability to be found outside the interval $\pm\Delta x$.

Fourier Series

An example of the first type of function is shown in Fig. 6.10: a periodic square wave of width a with a repeat distance (or spatial period) of λ . To keep the math simple, we will consider only even functions* like Fig. 6.10, that is, functions symmetric about the origin, such that $f(x) = f(-x)$. It turns out that any such periodic function can be expressed as

$$f(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{2\pi n}{\lambda} x\right) = \sum_{n=0}^{\infty} A_n \cos(k_n x) \quad (6.25)$$

where the A_n are constants, called the **Fourier coefficients** for the function $f(x)$. Equation (6.25) is called a **Fourier sum** or **Fourier series**. Notice that the different terms in the series correspond to different wavelengths. The first term ($n = 0$) is a constant term, and the succeeding terms have wavelengths

$$\lambda, \quad \frac{\lambda}{2}, \quad \frac{\lambda}{3}, \quad \dots, \quad \frac{\lambda}{n}, \quad \dots$$

The longest wavelength term ($n = 1$) is called the *fundamental* and the other terms are called *higher harmonics*. Corresponding to these wavelengths are the wave numbers

$$k_1 = \frac{2\pi}{\lambda}, \quad k_2 = 2 \cdot \frac{2\pi}{\lambda}, \quad k_3 = 3 \cdot \frac{2\pi}{\lambda}, \quad \dots \quad k_n = n \cdot \frac{2\pi}{\lambda}, \quad \dots$$

For any given function $f(x)$, the computation of the Fourier coefficients A_n is straightforward, but we won't need the details here. (See Problems 6.35 and 6.36.) For the square wave of Fig. 6.10, the coefficients turn out to be (Problem 6.32)

$$A_n = \frac{2}{\pi n} \sin\left(\frac{\pi a n}{\lambda}\right) \quad (6.26)$$

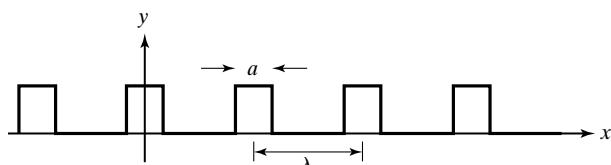
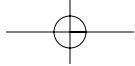
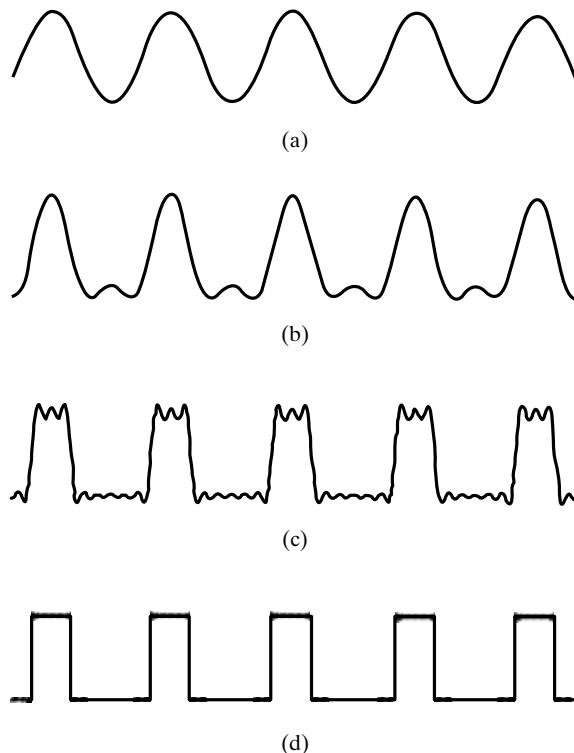


FIGURE 6.10

A simple periodic function, a series of square pulses with width a and repeat length λ .

*Odd periodic functions, such that $f(x) = -f(-x)$, can be written as a sum of sines. The general periodic function can be written as a sum of both sines and cosines.

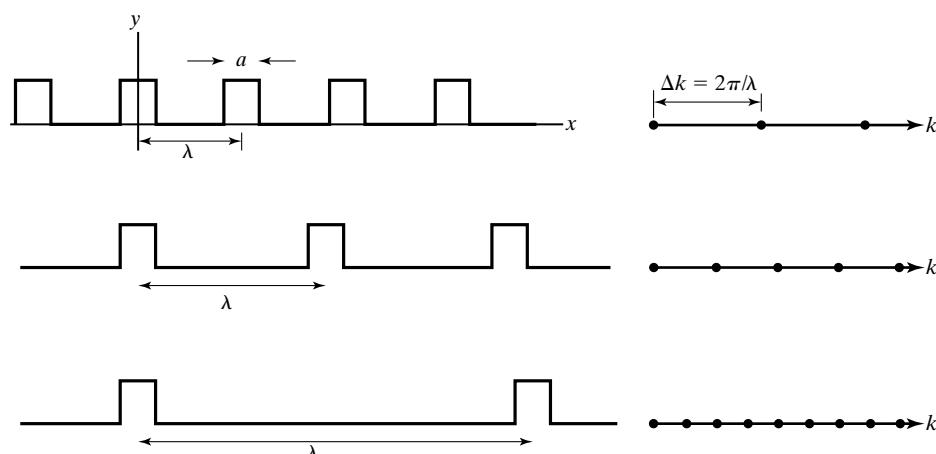
Section 6.7 • Wave Packets and Fourier Analysis **183****FIGURE 6.11**

The Fourier sum version of the periodic function in Fig. 6.10.
(a) The fundamental term only.
(b) The sum of the first 3 sinusoidal terms, **(c)** the first 8 terms, **(d)** the first 30 terms.

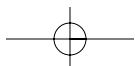
Although (6.25) is a sum with an infinite number of terms, in practice, the first several terms are often enough to approximate the function remarkably well, as illustrated in Fig. 6.11.

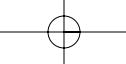
Fourier Integrals

Now let us consider the case of nonperiodic functions such as the wave packet of Fig. 6.9. A nonperiodic function can be regarded as a periodic function in the limit that the repeat distance goes to infinity, as in Fig. 6.12. Note that the spacing of the k 's in the Fourier sum (6.25) approaches zero as the repeat

**FIGURE 6.12**

A periodic function with repeat length λ becomes nonperiodic in the limit that $\lambda \rightarrow \infty$. As the repeat length λ grows, the spacing of k 's in the Fourier sum goes to zero.





184 Chapter 6 • Matter Waves

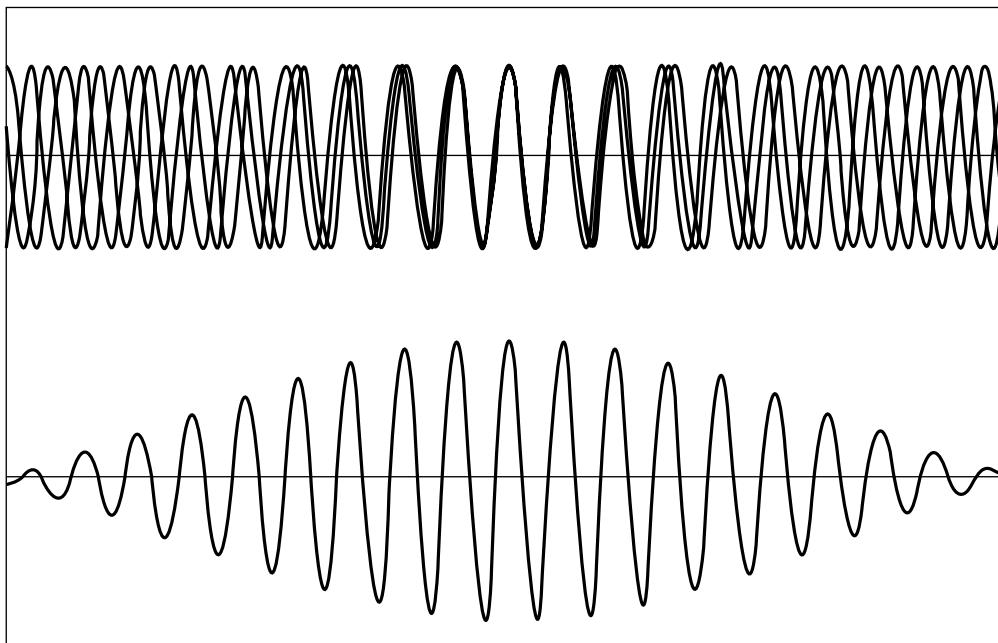


FIGURE 6.13

A wave packet (below) is made up of a series of sinusoidal waves of various wavelengths (above).

distance λ goes to infinity. (The spacing is $\Delta k = 2\pi/\lambda$, which approaches 0 as λ approaches ∞ .) Thus, while the periodic function is made up of a discrete sum of k 's, the nonperiodic function is made up of a continuous distribution of k 's and the Fourier sum becomes a **Fourier integral**

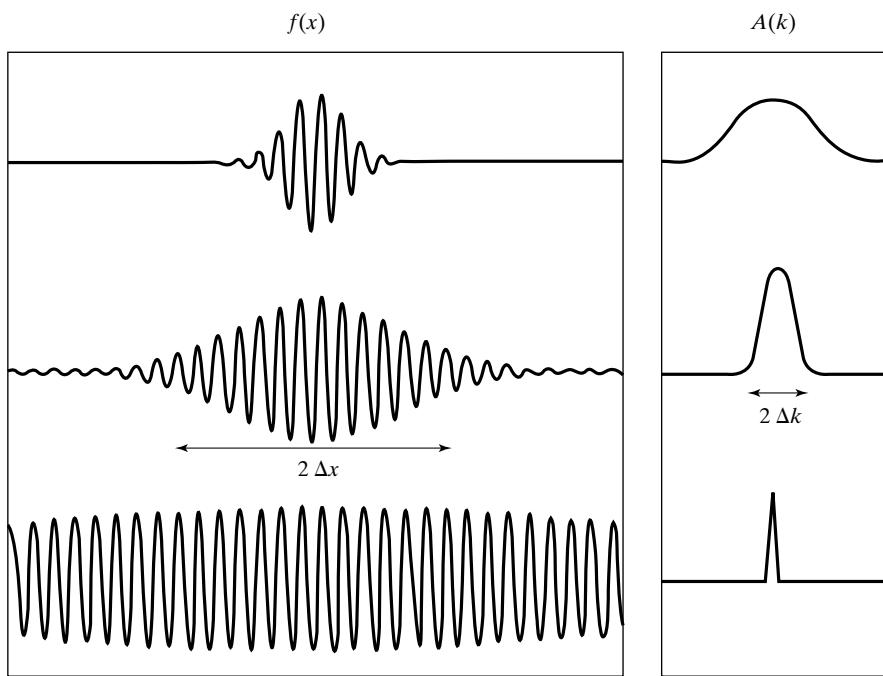
$$f(x) = \int A(k) \cos kx dk \quad (6.27)$$

The function $A(k)$ is called the **Fourier transform** of $f(x)$; it gives the distribution of wave numbers k that make up the wave packet. The way a wave packet is formed by adding up cosines with a distribution of wavelengths is illustrated in Fig. 6.13. We will now show that if $f(x)$ is a wave packet of size Δx , the wave packet is made up of a range of wave numbers Δk such that

$$\Delta x \Delta k \approx 1 \quad (6.28)$$

The constant on the right-hand side of (6.28) has order of magnitude 1, but its exact value depends on how we define the “widths” Δx and Δk . The relation (6.28) is illustrated in Fig. 6.14, which shows that a large Δx corresponds to a small Δk and vice versa: To make a very narrow wave packet (small Δx), you need a large spread in wavelengths, corresponding to a large spread in wave numbers (big Δk).

To establish (6.28), we continue to consider the special case of a wave packet that is an even function of x , made up of cosines only. We examine a wave packet, centered on the origin, consisting of wavelengths in the range $\lambda \pm \Delta\lambda$. We focus attention on just two terms: cosine terms with wavelengths $\lambda - \Delta\lambda$ and $\lambda + \Delta\lambda$, at the extreme ends of the range of wavelengths making up the packet (so their difference in wavelength is $2\Delta\lambda$). The spread $\Delta\lambda$

**FIGURE 6.14**

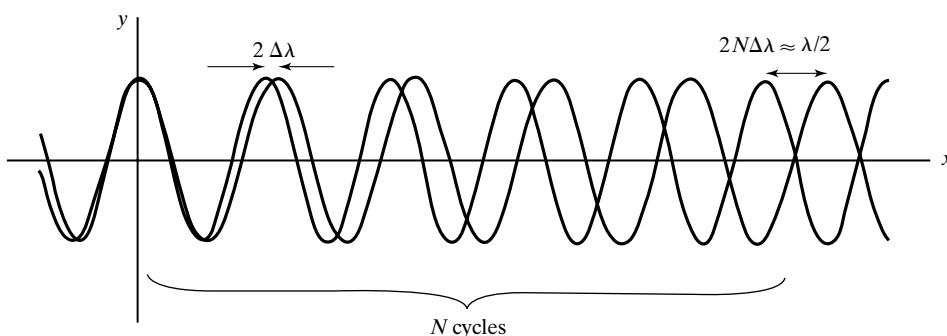
A narrow wave packet (small Δx) corresponds to a large spread of wavelengths (large Δk). A wide wave packet (large Δx) corresponds to a small spread of wavelengths (small Δk).

corresponds to a spread in wave numbers Δk as follows. Since $k = 2\pi/\lambda$, we have

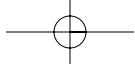
$$\Delta k = \left| \frac{dk}{d\lambda} \right| \Delta \lambda = \frac{2\pi}{\lambda^2} \Delta \lambda \quad (6.29)$$

(Here, the absolute value sign is necessary since $\Delta \lambda$ and Δk are positive by definition.)

Figure 6.15 shows these two cosine functions. Near $x = 0$, the two cosines are in phase and they add. However, away from $x = 0$, the phase difference between the two cosines grows progressively larger because of their wavelength difference. One cycle away from the origin, the peaks of the two cosines are a distance $2 \Delta \lambda$ apart. N cycles away from $x = 0$, the peaks of the two functions are $2N \Delta \lambda$ apart. When the two cosines are out of phase by 180° (that is, when $2N \Delta \lambda = \lambda/2$), the two waves cancel. This destructive interference occurs N wavelengths from the origin, approximately at position $x = N\lambda$.

**FIGURE 6.15**

Two cosine functions with wavelengths differing by $2 \Delta \lambda$.



186 Chapter 6 • Matter Waves

Thus, the half-width of the packet Δx is given by $N\lambda$ because the point $x = N\lambda$ is where the various waves have dephased sufficiently to sum to zero. We now have

$$2N \Delta\lambda = \frac{\lambda}{2} \quad \text{and} \quad \Delta x \approx N\lambda$$

Combining these equations to eliminate N , we get

$$\Delta x \approx \frac{\lambda^2}{4 \Delta\lambda} = \frac{1}{4} \frac{2\pi}{\Delta k}$$

or

$$\Delta x \Delta k \approx \pi/2$$

where we used (6.29) to relate $\Delta\lambda$ and Δk . Having established (6.28),* we can now understand this relation intuitively. The larger the spread in wavelengths, the more rapidly the waves dephase as one moves away from the center of the wave packet and the more rapidly the packet decays to zero: Bigger Δk means smaller Δx and vice versa.

We have rather loosely defined the spread, or **uncertainty**, (either Δx or Δk) as the half-width of the distribution. A more precise and frequently used definition is the root-mean-square, or rms, uncertainty, described in Sec. 3.8, and defined as

$$\Delta x = \sqrt{\langle (x - x_0)^2 \rangle} \quad (6.30)$$

where the brackets $\langle \dots \rangle$ indicate an appropriate average over all values of x and x_0 denotes the center of the wave packet. This rms uncertainty turns out to be somewhat smaller than the half-width uncertainty, as illustrated in Problems 6.33 and 6.34. When using the rms uncertainties, the relation (6.28) can be rigorously shown to be

$$\Delta x \Delta k \geq \frac{1}{2} \quad (6.31)$$

For uniformity, we will always use the form (6.31). This relation is sometimes called the wave packet uncertainty relation since it relates the spreads, or uncertainties, in x and k . Many authors write $\Delta x \Delta k \gtrsim 1$; in practice, this inequality is used mainly to give rough estimates of various quantities, and in such cases, the odd factor of 2 is of little consequence.

We have so far discussed a wave as a function of position x at a fixed time t . We can equally well consider the wave as a function of time t at a particular position x . Any wave pulse considered as a function of t can be expressed as a superposition of sinusoidal functions $\cos \omega t$ and $\sin \omega t$, where the angular frequency ω is related to the frequency f and period T as usual by $\omega = 2\pi f = 2\pi/T$. The analysis of such a function of time $g = g(t)$ is precisely as discussed above, but with x replaced by t and $k = 2\pi/\lambda$ replaced by $\omega = 2\pi/T$. The synthesis of a pulse localized in time requires a continuous

*That the right side has $\pi/2$ instead of 1 is irrelevant since this is just an order-of-magnitude argument.

spread $\Delta\omega$ of angular frequencies. The smaller the interval Δt in which the wave is localized, the larger must be the spread $\Delta\omega$ of angular frequencies. Specifically, the spreads Δt and $\Delta\omega$ satisfy an inequality exactly analogous to (6.31):

$$\Delta t \Delta\omega \geq \frac{1}{2} \quad (6.32)$$

In the next two sections we will see that the two inequalities (6.31) and (6.32) are especially important in the case of matter waves, for which they imply the celebrated Heisenberg uncertainty principle. In Section 6.8 we discuss the consequences of the inequality (6.31), and in Section 6.9 those of (6.32).

The inequality (6.32) has important implications for the transmission of information, as the following example illustrates.

Example 6.2

A TV picture, composed of 525 horizontal lines, refreshes 30 times a second. Hence, each line is drawn across the screen in $1/(30 \times 525) \text{ s} = 6.3 \times 10^{-5} \text{ s}$. What is the approximate range of frequencies Δf at which a TV transmitter must be able to broadcast, if the horizontal and vertical resolutions in the TV picture are to be about the same?

The vertical resolution of detail is $1/525$ of the screen height. Therefore, to achieve a comparable horizontal resolution, each individual line should be capable of becoming brighter or darker in about $1/525$ of a screen width. Since the beam travels across the screen in $6.3 \times 10^{-5} \text{ s}$, the transmitter must be able to send bright or dark pulses of total duration

$$2 \Delta t \approx \frac{1}{525} \times 6.3 \times 10^{-5} \text{ s} = 1.2 \times 10^{-7} \text{ s}$$

According to (6.32), this requires a spread of frequencies

$$\Delta\omega \geq \frac{1}{2 \Delta t} \approx 8.3 \times 10^6 \text{ s}^{-1}$$

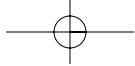
or

$$\Delta f = \frac{\Delta\omega}{2\pi} \gtrsim 1.3 \times 10^6 \text{ Hz}$$

The range of frequencies, $\pm\Delta f$, contained in a signal is called the *bandwidth*. Thus, we can say that TV video signals require a total bandwidth of at least* $2\Delta f = 2.6 \times 10^6 \text{ Hz}$. By comparison, high-fidelity sound systems require a frequency range of $4 \times 10^4 \text{ Hz}$ [†] in order to faithfully reproduce frequencies in the audio range of human hearing. The much higher bandwidth required for video signals delayed the development of videotape recorders many years after that of audiotape recorders.

*In fact, each television channel is assigned a bandwidth of 6 MHz, somewhat more than is needed for the standard analog color picture. Surprisingly, high-definition digital television (HDTV) uses the same bandwidth for each channel. Special digital image-compression techniques are used to squeeze the high resolution image into the relatively small bandwidth.

[†]The range of human hearing is 20 Hz to 20,000 Hz. It turns out that to accurately reproduce a 20 kHz wave-form, one needs frequencies up to twice that value, 40 kHz.



188 Chapter 6 • Matter Waves

To conclude this section, we note that one can apply the inequality (6.31) to the special case of a perfectly sinusoidal wave like $A \sin(kx - \omega t)$. This wave has an exactly defined wave number k and hence has $\Delta k = 0$. According to (6.31), if Δk approaches zero, Δx must approach ∞ , and this is exactly what we knew already: A perfectly sinusoidal wave must extend periodically through all of space and therefore has $\Delta x = \infty$. Of course, the values $\Delta k = 0$ and $\Delta x = \infty$ are an extremely special case of the inequality (6.31). Nevertheless, it is worth recognizing that (6.31) does cover even this extreme case.

6.8 The Uncertainty Relation for Position and Momentum

We have seen that the wave function of a single particle is spread out over some interval. This means that a measurement of the particle's position x may yield any value within this interval. (To simplify our discussion, we suppose that our particle moves in one dimension and so has a single coordinate x .) Therefore, the particle's position is *uncertain* by an amount $\pm \Delta x$, and we refer to Δx as the *uncertainty in the position*. As mentioned at the end of Section 6.4, the standard interpretation of quantum mechanics is that this uncertainty is not just a reflection of our ignorance of the particle's position. Rather, the particle *does not have a definite position*. The uncertainty exists in nature, not just in the mind of the physicist. The uncertainty Δx can be smaller in some states than in others; but for any given state, specified by a wave function $\Psi(x, t)$, there is some nonzero interval within which the particle may be found, and the particle's position is simply not defined any more precisely than that.

In Section 6.7 we saw that the wave function that describes a particle can be built up from sinusoidal waves, but that this requires a spread of different wave numbers k (or wavelengths λ). From the de Broglie relation,

$$p = \frac{h}{\lambda} = \hbar k$$

it follows that a spread of wave numbers implies a spread of momenta; that is, the particle's momentum p , like its position x , is uncertain. A measurement of the momentum may yield any of several values in a range given by

$$\Delta p = \hbar \Delta k \quad (6.33)$$

We have seen that the spreads Δx and Δk are not independent, but always satisfy the inequality (6.31),

$$\Delta x \Delta k \geq \frac{1}{2}$$

If we multiply this relation by \hbar , we find that

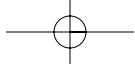
$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad (6.34)$$

This is one of several inequalities called the **Heisenberg uncertainty relations** and known collectively as the **Heisenberg uncertainty principle**. It implies that both the position and momentum of a particle have uncertainties in the sense just described. One can find states for which Δx is small, but (6.34) tells us that Δp will be large; one can also find states for which Δp is small, but Δx will be large. In all cases their product, $\Delta x \Delta p$, will never be less than $\hbar/2$.

Werner Heisenberg
(1901–1976, German)



After earning his PhD at Munich, Heisenberg worked with Born and then Bohr. His many contributions to modern physics include an early formulation of quantum mechanics in terms of matrices, several ideas in nuclear physics, and the famous uncertainty principle, for which he won the 1932 Nobel Prize in physics. He remained in Germany during World War II and worked on nuclear reactor design. The possibility that he might be working on an atomic bomb for the Nazis so frightened the Allies that a plan was devised to have him assassinated. An American agent, named Moe Berg, posed as a physicist and met with Heisenberg when he was visiting neutral Switzerland in late 1944. After talking with Heisenberg, agent Berg decided that the Germans had made little progress toward a bomb and chose not to kill him.

Section 6.8 • The Uncertainty Relation for Position and Momentum **189**

In classical physics it was taken for granted that particles have definite values of their position x and momentum p . It was recognized, of course, that x and p could not be measured with perfect accuracy. But it was assumed that with enough care, one could make both experimental uncertainties as small as one pleased. Heisenberg's uncertainty relation (6.34) shows that these assumptions were incorrect. There are intrinsic uncertainties, or spreads, Δx and Δp in the position and momentum of any particle. Whereas either one of Δx and Δp can be made as small as one pleases, their product can never be less than $\hbar/2$.

We now know that the uncertainty principle applies to all particles. On the macroscopic level, however, it is seldom important, as the following example illustrates.

Example 6.3

The position x of a 0.01-g pellet has been carefully measured and is known within $\pm 0.5 \mu\text{m}$. According to the uncertainty principle, what are the minimum uncertainties in its momentum and velocity, consistent with our knowledge of x ?

If x is known within $\pm 0.5 \mu\text{m}$, the spread $\pm \Delta x$ in the position is certainly no larger than $0.5 \mu\text{m}$:

$$\Delta x \leq 0.5 \mu\text{m}$$

According to the uncertainty relation (6.34), this implies that the momentum is uncertain by an amount

$$\Delta p \geq \frac{\hbar}{2 \Delta x} \geq \frac{10^{-34} \text{ J} \cdot \text{s}}{10^{-6} \text{ m}} = 10^{-28} \text{ kg} \cdot \text{m/s}$$

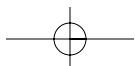
Therefore, the velocity $v = p/m$ is uncertain by*

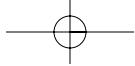
$$\Delta v = \frac{\Delta p}{m} \geq \frac{10^{-28} \text{ kg} \cdot \text{m/s}}{10^{-5} \text{ kg}} = 10^{-23} \text{ m/s}$$

Clearly, the inevitable uncertainties in p and v required by the uncertainty principle are of no practical importance in this case. (To appreciate how small 10^{-23} m/s is, notice that at this speed our pellet would take about a million years to cross an atomic diameter.)

Although the uncertainty principle is seldom important on the macroscopic level, it is frequently very important on the microscopic level, as the next example illustrates.

*The mass of a stable particle has no uncertainty, so we can treat m as a constant in the relation $v = p/m$.





190 Chapter 6 • Matter Waves

Example 6.4

An electron is known to be somewhere in an interval of total width $a \approx 0.1 \text{ nm}$ (the size of a small atom). What is the minimum uncertainty in its velocity, consistent with this knowledge?

If we know the electron is certainly inside an interval of total width a ,

$$\Delta x \leq \frac{a}{2} \quad (6.35)$$

(Remember that Δx is the spread from the central value out to either side.) According to the uncertainty relation (6.34), this implies that

$$\Delta p \geq \frac{\hbar}{2 \Delta x} \geq \frac{\hbar}{a} \quad (6.36)$$

This implies that $\Delta v = \Delta p/m \gtrsim \hbar/(am)$ or

$$\Delta v \geq \frac{\hbar c^2}{amc^2} = \frac{200 \text{ eV} \cdot \text{nm}}{(0.1 \text{ nm}) \times (0.5 \times 10^6 \text{ eV})} c = \frac{c}{250} \approx 10^6 \text{ m/s}$$

(where we multiplied numerator and denominator by c^2 to take advantage of the useful combinations $\hbar c$ and mc^2). This large uncertainty in v shows the great importance of the uncertainty principle for systems with atomic dimensions.

Perhaps the most dramatic consequence of the uncertainty principle is that a particle confined in a small region cannot be exactly at rest, since if it were, its momentum would be precisely zero, which would violate (6.36). Since its momentum cannot be precisely zero, the same is true of its kinetic energy. Therefore, the particle has a minimum kinetic energy, which we can estimate as follows: Since the momentum is spread out by an amount given by (6.36) as

$$\Delta p \geq \frac{\hbar}{a} \quad (6.37)$$

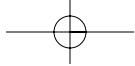
the magnitude of p must be, on average, at least of this same order. Thus the kinetic energy, whether it has a definite value or not, must on average have magnitude

$$\langle K \rangle = \left\langle \frac{p^2}{2m} \right\rangle \gtrsim \frac{(\Delta p)^2}{2m} \quad (6.38)$$

or, by (6.37)

$$\langle K \rangle \gtrsim \frac{\hbar^2}{2ma^2} \quad (6.39)$$

The energy (6.39) is called the **zero-point energy**. It is the minimum possible kinetic energy for a quantum particle confined inside a region of

Section 6.8 • The Uncertainty Relation for Position and Momentum **191**

width a . The kinetic energy can, of course, be larger than this, but it cannot be any smaller.

Example 6.5

What is the minimum kinetic energy of an electron confined in a region of width $a \approx 0.1$ nm, the size of a small atom?

According to (6.39),

$$\langle K \rangle \gtrsim \frac{\hbar^2}{2ma^2} = \frac{(\hbar c)^2}{(2mc^2)a^2} = \frac{(200 \text{ eV} \cdot \text{nm})^2}{(10^6 \text{ eV}) \times (0.1 \text{ nm})^2} = 4 \text{ eV}$$

This lower bound is satisfactorily consistent with the known kinetic energy, 13.6 V, of an electron in the ground state of a hydrogen atom.*

The bound (6.39) gives a useful estimate of the minimum kinetic energy of several other systems. For more examples, see Problems 6.39, 6.42, and 6.43.

We have so far written the uncertainty relation only for the case of a particle moving in one dimension. In three dimensions there is a corresponding inequality for each dimension separately:

$$\Delta x \Delta p_x \geq \frac{\hbar}{2} \quad \Delta y \Delta p_y \geq \frac{\hbar}{2} \quad \Delta z \Delta p_z \geq \frac{\hbar}{2} \quad (6.40)$$

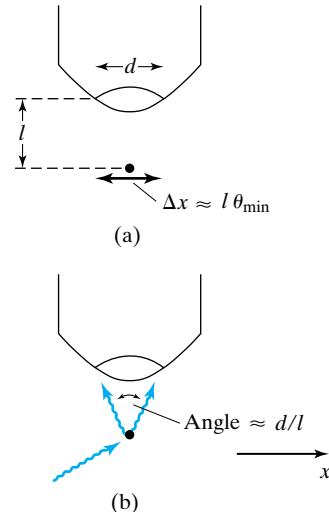
where x, y, z are the particle's three coordinates and p_x, p_y, p_z , the three components of its momentum. (See Problem 6.42 for an application.)

Heisenberg's Microscope

The uncertainty principle can be illustrated by several thought experiments, the best known of which is sometimes called the Heisenberg microscope. In this thought experiment a classical physicist — reluctant to accept the uncertainty principle — tries to disprove it by showing that he can measure the position and momentum of a particle with uncertainties smaller than are allowed by the uncertainty relation (6.34).

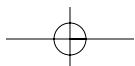
To find the position x of the particle, our classical physicist observes it with a microscope, as shown in Fig. 6.16. Now, it is a fact — well known in classical physics — that the resolution of any microscope is limited by the diffraction of light. Specifically, the angular resolution θ_{\min} , (the minimum angle at which two points can be told apart) is given by the so-called Rayleigh criterion,

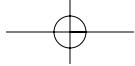
$$\theta_{\min} \approx \frac{\lambda}{d} \quad (6.41)$$

**FIGURE 6.16**

The Heisenberg microscope.
(a) The minimum experimental uncertainty in the particle's position x is determined by the microscope's resolution as $\Delta x \approx l \theta_{\min} \approx l\lambda/d$. **(b)** The direction of a photon entering the microscope is uncertain by an angle of order d/l ; therefore, the photon gives the particle a momentum (in the x direction) which is uncertain by $\Delta p_x \approx p_\gamma d/l$.

*Recall that we saw in Sec. 5.6 that the kinetic energy is the negative of the total energy ($E = -13.6$ eV). Also, here we have treated an electron in one dimension. If one uses the inequalities (6.40) to include the motion in all three dimensions, one finds $\langle K \rangle \gtrsim 12$ eV, in excellent agreement with the observed 13.6 eV.





192 Chapter 6 • Matter Waves

where λ is the wavelength of light used and d the diameter of the objective lens. If the particle is a distance l below the lens, the minimum uncertainty in x is [see Fig. 6.16(a)]

$$\Delta x \approx l\theta_{\min} \approx \frac{l\lambda}{d} \quad (6.42)$$

(where we assume for simplicity that all angles are small, so that $\sin \theta \approx \theta$). Our classical physicist is aware of this limitation, but points out that he can make Δx as small as he pleases, for example, by using light of very short wavelength λ .

Simply to pin down the particle's position with arbitrarily small Δx does not itself conflict with the uncertainty principle. Our classical physicist must show that he can also know the momentum with a suitably small uncertainty; and if we recall that light is quantized, we can quickly show that this is impossible: In order to observe the particle, he must allow at least one photon to strike it, and this collision will change the particle's momentum. He has no way of knowing which part of the lens the photon passed through since the lens sends *any* light from the object through the same image point. Therefore, the direction in which the photon approached the lens is uncertain by an angle of order d/l [See Fig. 6.16(b).] This means that the x component of the photon's momentum is uncertain by an amount of order $p_y d/l$. Since the particle was struck by the photon, the x component of the particle's momentum is now uncertain by at least this same amount; that is,

$$\Delta p_x \gtrsim p_y \frac{d}{l} = \frac{\hbar}{\lambda} \cdot \frac{d}{l} \quad (6.43)$$

Our classical physicist can make this uncertainty in p_x as small as he pleases, for example by making λ large. But comparing (6.42) and (6.43), we see that whatever he does to reduce Δp_x will increase Δx and vice versa. In particular, multiplying (6.42) by (6.43), we find that

$$\Delta x \Delta p_x \gtrsim \hbar \quad (6.44)$$

and our classical physicist has failed in his attempt to disprove the uncertainty principle.*

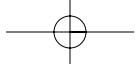
The uncertainty principle is a general result that follows from the particle-wave duality of nature. We should emphasize that our analysis of the Heisenberg microscope is not an alternative proof of this general result; it serves only to illustrate the inevitable appearance of the uncertainty principle in the context of one particular experiment.

6.9 The Uncertainty Relation for Time and Energy

Just as the inequality $\Delta x \Delta k \geq \frac{1}{2}$ implies the position-momentum uncertainty relation, $\Delta x \Delta p \geq \hbar/2$, so the inequality (6.32)

$$\Delta t \Delta \omega \geq \frac{1}{2} \quad (6.45)$$

*The fact that we have found $\Delta x \Delta p_x \gtrsim \hbar$, rather than $\hbar/2$, is not significant, since the arguments leading to (6.44) were only order-of-magnitude arguments.

Section 6.9 • The Uncertainty Relation for Time and Energy **193**

implies a corresponding relation for time and energy. Specifically, if we multiply by \hbar , we find the **time-energy uncertainty relation**

$$\Delta t \Delta E \geq \frac{\hbar}{2} \quad (6.46)$$

Here ΔE is the uncertainty in the particle's energy: A quantum particle generally does not have a definite energy, and measurement of its energy can yield any answer within a range $\pm \Delta E$. To understand the significance of Δt , recall that the inequality (6.45) arose when we considered a wave pulse as a function of time t at one fixed position x . The time Δt characterizes the time spent by the pulse at that position. Thus, for a quantum wave, Δt characterizes the time for which the particle is likely to be found at the position x . According to (6.46), if Δt is small, the particle must have a large uncertainty ΔE in its energy and vice versa.

If a particle has a definite energy, then $\Delta E = 0$, and (6.46) tells us that Δt must be infinite. That is, a quantum particle with definite energy stays localized in the same region (and in the same state, in fact) *for all time*. States with this property are the quantum analog of Bohr's stationary orbits and are called stationary states, as we discuss in Chapter 7.

If a particle (or, more generally, any quantum system) does *not* remain in the same state forever, Δt is finite and (6.46) tells us that ΔE cannot be zero; that is, the energy must be uncertain. For example, any unstable state of an atom or nucleus lives for a certain finite time Δt , after which it decays by emitting a particle (an electron, photon, or α particle, for example). This means that the energy of any unstable atom or nucleus has a minimum uncertainty*

$$\Delta E \approx \frac{\hbar}{2 \Delta t} \quad (6.47)$$

Since the energy of the original unstable state is uncertain, the same is true of the ejected particle. In some cases one can measure both the spread of energies of the ejected particles (from many decays of identical unstable systems) and the lifetime Δt ; one can then confirm the relation (6.47). In many applications one measures one of the quantities ΔE or Δt , then uses (6.47) to estimate the other.

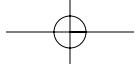
Example 6.6

Many excited states of atoms are unstable and decay by emission of a photon in a time of order $\Delta t \approx 10^{-8}$ s. What is the minimum uncertainty in the energy of such an atomic state?

According to (6.47), the minimum uncertainty in energy is

$$\Delta E \approx \frac{\hbar}{2 \Delta t} = \frac{\hbar c}{2c \Delta t} \approx \frac{200 \text{ eV} \cdot \text{nm}}{2 \times (3 \times 10^{17} \text{ nm/s}) \times (10^{-8} \text{ s})} \approx 3 \times 10^{-8} \text{ eV}$$

*In (6.47) we have used the symbol \approx because several different definitions of ΔE and Δt are commonly used. For example, Δt can be defined as the half-life (discussed in Section 1.9) or the mean life (to be discussed in Chapter 17), and the precise form of the relation depends on which definition we adopt. For the case of an unstable particle, (6.47) is exact if we take ΔE to be the so-called half-width at half-height and Δt to be the mean life.



194 Chapter 6 • Matter Waves

Compared to the several eV between typical atomic energy levels, this uncertainty ΔE is very small. Nevertheless, the resulting spread in the energy, and hence frequency, of the ejected photon is easily measurable with a modern spectrometer.

Nowadays, the frequencies of photons ejected in atomic transitions are used as standards for the definition and calibration of frequency and time. Because of the uncertainty principle, the frequency of any such photon is uncertain by an amount

$$\Delta\omega = \frac{\Delta E}{\hbar} \approx \frac{1}{2\Delta t}$$

where Δt is the lifetime of the emitting state. Therefore, it is important to choose atomic states with very long lifetimes Δt to use as standards.

6.10 Velocity of a Wave Packet*

*In this section we discuss a puzzle concerning the speed of matter waves. This is one of those curious problems that are important in principle, but relatively unimportant in practice. In particular, this material will not be needed again in this book, and you can therefore omit this section if you wish.

The problem that we address in this section is this: When we compute the speed of a matter wave, we find that the wave's speed is not equal to the speed of the particle that the wave describes. To see this, consider a single particle of mass m moving nonrelativistically with momentum p , in a region free of all forces. The particle's energy is just its kinetic energy:

$$E = K = \frac{p^2}{2m}$$

From the de Broglie relations (6.1), we can find the frequency and wavelength:

$$f = \frac{E}{\hbar} = \frac{p^2}{2mh} \quad \text{and} \quad \lambda = \frac{h}{p}$$

For any wave, the wave speed is just the frequency times the wavelength. Thus, for our matter wave

$$v_{\text{wave}} = f\lambda = \frac{p^2}{2mh} \frac{h}{p} = \frac{p}{2m} \quad (6.48)$$

However, for a nonrelativistic particle with velocity v_{part} , we know that $p = mv_{\text{part}}$ and hence $v_{\text{part}} = p/m$. Thus (6.48) implies that

$$v_{\text{wave}} = \frac{v_{\text{part}}}{2} \quad (6.49)$$

At first sight this result is very surprising, since it is natural to assume that the velocity v_{wave} of a matter wave should be the same as the velocity v_{part} of the particle to which it corresponds.

Section 6.10 • Velocity of a Wave Packet 195

To understand why (6.49) is perfectly acceptable (and correct), we must consider carefully the significance of the wave velocity, namely, that v_{wave} is the velocity with which any crest of a sinusoidal wave moves, as was shown in Fig. 6.8. We argued in Section 6.7 that a sinusoidal wave is an idealization that cannot occur in practice. The wave representing a particle is necessarily a wave *packet* like that in Fig. 6.9, and as we will see shortly, the velocity v_{pack} with which a wave packet moves as a whole is not necessarily the same as the velocity, v_{wave} , of its individual crests. Since the particle is represented by the whole wave packet, rather than any single wave crest, it is v_{pack} that should equal the velocity of the corresponding particle. We will prove in a moment that indeed $v_{\text{pack}} = v_{\text{part}}$, but let us first consider how it is that v_{pack} and v_{wave} can be different.

To understand the difference between v_{pack} and v_{wave} , consider Fig. 6.17, which shows the observed behavior of a wave packet on deep water. We see that the packet as a whole is moving to the right at a constant velocity v_{pack} (often called the **group velocity**). On closer inspection, we see that any given wave crest, like the one carrying the surfer, is traveling to the right faster than the packet as a whole. In the first picture the surfer's crest is at the center of the packet, but in each successive picture it is closer to the front of the packet, until in the sixth picture the surfer's crest disappears entirely and the surfer sinks ignominiously into the water. As a packet of this type advances, wave crests steadily move forward and disappear at the front of the packet, while others appear at the rear, allowing the packet to maintain its overall shape. The behavior of the packet in Fig. 6.17 is typical of waves on deep water and can, in fact, be observed by carefully watching the bow waves of a boat on a calm lake.*

To understand mathematically how the behavior shown in Fig. 6.17 occurs, we must recognize that a wave packet is the result of the interference of many different sinusoidal waves, as discussed in Section 6.7. When one superposes sinusoidal waves of different frequencies, there are two main possibilities: For some waves (light waves in vacuum, for example) the wave speed, v_{wave} , is the same for all frequencies. This means that the various sine waves all travel at the same speed, and their interference pattern (that is, the packet) is carried along at the same speed as the individual waves. In this case $v_{\text{pack}} = v_{\text{wave}}$.

For many waves, including waves on deep water and matter waves, the wave speed, v_{wave} , is different for different frequencies. In this case the various sine waves that make up the packet move at different speeds, and the resulting interference pattern shifts steadily, relative to the component sine waves. This means that the interference pattern (or wave packet) moves at a speed different from that of the component sine waves: $v_{\text{pack}} \neq v_{\text{wave}}$.

A realistic wave packet, like that of Fig. 6.17, is a superposition of infinitely many different sine waves, and the mathematical analysis of such a wave is beyond our scope here. Instead, we consider a superposition of just two sine waves. Although this does not produce a realistic wave packet, it illustrates the main ideas and gives us the correct expression for the velocity, v_{pack} , of the wave packet.

Let us consider, then, the superposition of two sine waves with equal amplitudes but with wave numbers that differ by Δk and frequencies that differ by $\Delta\omega$:

$$\Psi(x, t) = A \sin[(k + \Delta k)x - (\omega + \Delta\omega)t] + A \sin[kx - \omega t] \quad (6.50)$$

*On shallow water it turns out that $v_{\text{wave}} = v_{\text{pack}}$ (Problem 6.50), so a surfer can ride an individual crest until it breaks as it nears the shore.

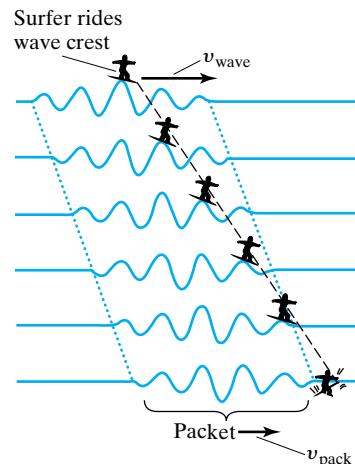
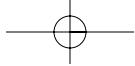


FIGURE 6.17

Six successive views of a wave packet, or group, moving to the right. The velocity of the whole packet is v_{pack} , often called the **group velocity**. The velocity of an individual crest, like the one carrying the surfer, is v_{wave} , the **wave velocity** (or phase velocity). For this wave, $v_{\text{wave}} > v_{\text{pack}}$ and the surfer moves steadily toward the front of the packet. The two sloping dotted lines indicate the motion of the front and back of the packet; the sloping dashed line indicates the motion of a single wave crest.



196 Chapter 6 • Matter Waves

Using the identity (Appendix B)

$$\sin \theta + \sin \phi = 2 \sin \frac{\theta + \phi}{2} \cos \frac{\theta - \phi}{2}$$

we can rewrite Ψ as (Problem 6.52)

$$\Psi(x, t) = 2A \sin(\bar{k}x - \bar{\omega}t) \cos\left(\frac{\Delta k}{2}x - \frac{\Delta \omega}{2}t\right) \quad (6.51)$$

where \bar{k} and $\bar{\omega}$ denote the mean wave number and mean frequency ($\bar{k} = k + \Delta k/2$ and $\bar{\omega} = \omega + \Delta \omega/2$). The important property of (6.51) is that it is the product of two terms. The first term is a sine wave with wave number and frequency corresponding to the average of the original two waves; the second, cosine, term has wave number and frequency corresponding to half the difference of the original two waves. If the original two waves were close together in frequency ($\Delta\omega$ and Δk small), this second term oscillates much more slowly than the first and forms the envelope for the more rapid oscillations of the first term, as in Fig. 6.18(b).

Since the individual crests in Fig. 6.18(b) are defined by the rapidly oscillating first term in (6.51), they travel to the right with speed (Problem 6.27)

$$v_{\text{wave}} = \bar{\omega}/\bar{k} \quad (6.52)$$

When $\Delta\omega$ and Δk are small, this is close to the wave speed of either of the original two waves. To find the velocity of the interference pattern, we must look at the second term in (6.51), which travels to the right with speed (again see Problem 6.27)

$$v_{\text{pack}} = \frac{\Delta\omega/2}{\Delta k/2} \approx \frac{d\omega}{dk}. \quad (6.53)$$

We can relate ω to k as follows: $\omega = 2\pi f = 2\pi(v_{\text{wave}}/\lambda) = v_{\text{wave}}k$. Thus

$$\omega = v_{\text{wave}}k \quad (6.54)$$

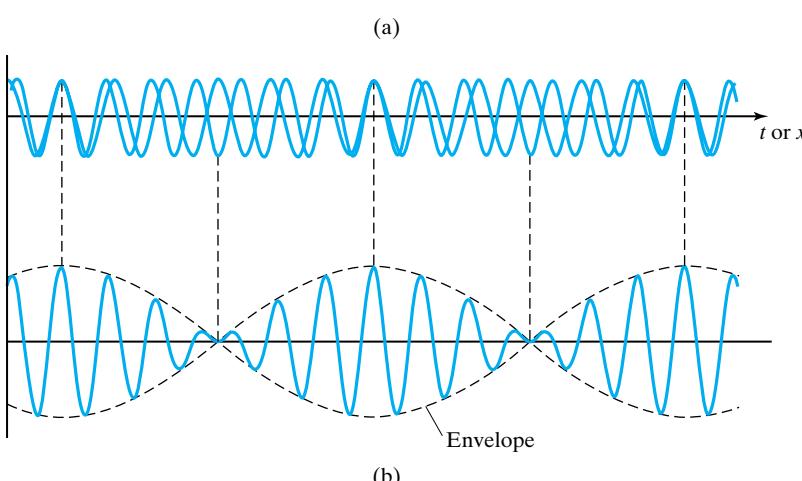


FIGURE 6.18

(a) If two waves with slightly different frequencies ω_1 and ω_2 are superposed, they are alternately in and out of step. (b) The resultant wave shows the phenomenon of beats, in which a wave of frequency $\bar{\omega} = (\omega_1 + \omega_2)/2$ is modulated by an envelope, which oscillates at the difference frequency.

If the wave speed is the same for all frequencies, v_{wave} is a constant, and differentiation of (6.54) gives

$$v_{\text{pack}} = \frac{d\omega}{dk} = v_{\text{wave}} \quad [\text{if } v_{\text{wave}} = \text{constant}]$$

This result agrees with our previous argument that if v_{wave} is the same for all frequencies, the interference pattern will be carried along at the speed v_{wave} . If, however, v_{wave} is not constant, differentiation of (6.54) yields two terms (v_{wave} plus $k \, dv_{\text{wave}}/dk$). Therefore, $d\omega/dk$ is not the same as v_{wave} , and the speed with which the envelope moves is different from v_{wave} .

The results just derived were for the superposition of two sine waves. As long as the range of different frequencies is small (as is usually the case), similar results can be proved for superpositions of any number of sine waves. In particular, for a wave packet like that in Fig. 6.17, one can prove that the individual crests move with velocity $v_{\text{wave}} = \bar{\omega}/\bar{k}$, where $\bar{\omega}$ and \bar{k} are the average frequency and wave number, but that the whole packet moves with the group velocity

$$v_{\text{pack}} = \frac{d\omega}{dk} \quad (6.55)$$

Only when v_{wave} is independent of frequency are v_{pack} and v_{wave} equal; otherwise, they are not.

We can now apply these ideas to matter waves, whose frequency and wave number are determined by the de Broglie relations

$$E = \hbar\omega \quad \text{and} \quad p = \hbar k$$

In this case (6.55) gives for the group velocity, v_{pack} ,

$$v_{\text{pack}} = \frac{d\omega}{dk} = \frac{dE}{dp} \quad (6.56)$$

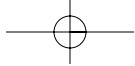
For a nonrelativistic particle moving freely, $E = p^2/2m$ and (6.56) implies that

$$v_{\text{pack}} = \frac{d}{dp} \left(\frac{p^2}{2m} \right) = \frac{p}{m} = v_{\text{part}}$$

That is, the velocity of the wave packet is the same as the particle's velocity, and the de Broglie wave packet does indeed move with the speed of the particle it represents.

CHECKLIST FOR CHAPTER 6

Concept	Details
De Broglie relations	$E = hf$ and $p = h/\lambda$ (6.1)
De Broglie's explanation of quantization of angular momentum	Waves in a confined region \Rightarrow quantization of wavelengths \Rightarrow quantization of angular momentum. (Sec 6.2)
Davisson-Germer experiment	First evidence for matter waves. (Sec. 6.3)
Probabilistic meaning of the quantum wave function	$ \Psi ^2 = \text{probability density}$ (6.15)
Parameters of a sinusoidal wave	$\lambda = 2\pi/k$ (6.20), $f = \omega/2\pi = 1/T$ (6.21), $v = \lambda f$ (6.22)



198 Chapter 6 • Matter Waves

De Broglie relations in terms of ω and k	$E = \hbar\omega$ (6.23) and $p = \hbar k$ (6.24)
Wave packet	A wave that is zero (or vanishingly small) outside some finite region
Fourier series	$f(x) = \sum A_n \cos nkx$ (6.25)
Fourier integral	$f(x) = \int A(k) \cos kx dk$ (6.27)
Wave packet uncertainty relations	$\Delta x \Delta k \geq 1/2$ (6.31) and $\Delta t \Delta \omega \geq 1/2$ (6.32)
Heisenberg uncertainty relations	$\Delta x \Delta p \geq \hbar/2$ (6.34) and $\Delta t \Delta E \geq \hbar/2$ (6.46)
Minimum kinetic energy of a confined particle	$\langle K \rangle \gtrsim \hbar^2/2ma^2$ (6.39)
Group velocity	$v_{\text{pack}} = d\omega/dk$ (6.55)

PROBLEMS FOR CHAPTER 6

SECTION 6.2 and 6.3 (De Broglie's Hypothesis and Experimental Verification)

- 6.1** • Use the de Broglie relation $\lambda = h/p$ to find the wavelength of a raindrop with mass $m = 1$ milligram and speed 1 cm/s. Does it seem likely that the wave properties of a raindrop could be easily detected?
- 6.2** • Use the de Broglie relation $\lambda = h/p$ to find the wavelength of a golf ball of mass 60 grams with speed 30 m/s. Does it seem likely that the wave properties of a golf ball could be easily detected?
- 6.3** • Use the de Broglie relation $\lambda = h/p$ to find the wavelength of electrons with kinetic energy 500 eV.
- 6.4** • Find the kinetic energy of an electron with the same wavelength as blue light ($\lambda \approx 450$ nm).
- 6.5** • Find the kinetic energy of a neutron with the same wavelength as blue light ($\lambda \approx 450$ nm).
- 6.6** • Compare the wavelengths of electrons and neutrons, both with $K = 3$ eV.
- 6.7** • Find and compare the wavelengths of an electron and a muon ($m_\mu = 207 m_e$) each with kinetic energy 15 keV.
- 6.8** • In order to investigate a certain crystal, we need a wave with $\lambda = 0.05$ nm. If we wish to use neutrons, what should be their kinetic energy? What if we use electrons? What for photons?
- 6.9** • Find the wavelength of an electron with energy $E = 2$ MeV. [Hint: The de Broglie relation $\lambda = h/p$ is correct at all energies, but since this energy is relativistic, you will have to use the relation $E^2 = (pc)^2 + (mc^2)^2$ to find p .]
- 6.10** • Find the wavelength of an electron with *kinetic* energy $K = 2$ MeV. (Note that here we are giving the kinetic energy, not the total energy. See the hint in Problem 6.9.)

6.11 •• Using the appropriate relativistic relations between energy and momentum, find and compare the wavelengths of electrons and photons at the three different kinetic energies: 1 keV, 1 MeV, 1 GeV.

6.12 •• (a) Use the relativistic relation between E and p to show that electrons and photons with the same energy E have different wavelengths. (Note: Even at relativistic energies the de Broglie relation $\lambda = h/p$ is correct.) (b) Show that their wavelengths approach equality as their common energy E gets much larger than $m_e c^2$.

6.13 •• At what common energy E do the wavelengths of electrons and photons differ by a factor of (a) 2, (b) 1.1, (c) 1.01? (See Problem 6.12.)

SECTION 6.4 and 6.5 (The Quantum Wave Function and Which Slit does the Electron Go Through?)

- 6.14** • Electrons with $K = 100$ eV are directed at two narrow slits a distance d apart. If the angle between the central maximum of the resulting interference pattern and the next maximum is to be 1° , what should d be?
- 6.15** • An experimenter wishes to arrange a two-slit experiment with 3-eV electrons so that the $n = 1$ maximum occurs at 15° . What will his slit separation, d , have to be?
- 6.16** • What are the dimensions of a matter wave $\Psi(\mathbf{r}, t)$ describing an electron in three dimensions? What are its SI units? [Hint: Probabilities are dimensionless]. (Why?)
- 6.17** •• A classical physicist is determined to find out which slit each electron passes through in the two-slit experiment (without disrupting the interference pattern). To this end, he places a molecule near one slit in

the hope that electrons passing through this slit will excite the molecule, causing it to give out a characteristic pulse of light. Show that this arrangement fares no better than the thought experiment using light that was described in Section 6.5.

SECTION 6.6 (Sinusoidal Waves)

- 6.18** • A wave described by (6.19) has $k = 6 \text{ rad/m}$ and $\omega = 22 \text{ rad/s}$. Find λ , f , and v .
- 6.19** • A traveling wave is given by $y(x, t) = A \sin(kx - \omega t)$ with $A = 4 \text{ cm}$, $k = 12 \text{ rad/cm}$, and $\omega = 2 \times 10^3 \text{ rad/s}$. Find the wave speed v , wavelength λ , and frequency f .
- 6.20** • What are the SI units of λ , T , f , k , ω , and v ?
- 6.21** • For green light ($\lambda \approx 550 \text{ nm}$), find k and ω .
- 6.22** • Find k and ω for X-rays with $\lambda = 0.05 \text{ nm}$.
- 6.23** • Use the de Broglie relation (6.24) to find λ and k for electrons with kinetic energy 300 eV.
- 6.24** • If we observe a point on a string with a fixed value of x , it will oscillate up and down as the wave (6.18) travels past it. Show that it oscillates with simple harmonic motion of frequency $f = 1/T$.
- 6.25** • (a) At any fixed point $x = x_0$, the traveling wave $y(x, t) = A \sin(kx - \omega t)$ can be expressed as $y(x_0, t) = A \sin(\omega t + \phi)$. Find ϕ . (b) For $x_0 = 0$, what is ϕ ? (c) By how much must one change x_0 [from its value in part (b)] so that ϕ is π larger than in part (b)? Express your answer in terms of λ .
- 6.26** •• (a) Prove that a crest of the wave (6.18) moves with speed $v = \lambda/T$ to the right. [Hint: Focus attention on one wave crest P (for example, the crest for which the argument of the sine function is $\pi/2$) in order to find an expression for x_p , in terms of t .] (b) Show that if the minus sign in (6.18) is replaced by a plus sign, the wave moves to the left.
- 6.27** •• Show that the crests of a wave $\Psi = A \sin(ax - bt)$ move to the right at a speed v equal to the ratio of the coefficients of t and x : $v = b/a$. (See the hint to Problem 6.26.)

SECTION 6.7 (Wave Packets and Fourier Analysis)

- 6.28** • A telephone line can transmit a range of frequencies $\Delta f \approx 2500 \text{ Hz}$. Roughly, what is the duration of the shortest pulse that can be sent over this line?
- 6.29** • A space probe sends a picture containing 500×500 elements, each containing a brightness scale with 256 possible levels. This scale requires eight binary digits. Thus, altogether $8 \times 500 \times 500 = 2 \times 10^6$ pulses are required to encode the picture for transmission. Suppose that the transmitter uses a bandwidth of 1000 Hz. (For the faint signals from distant space probes the bandwidth must be kept small to reduce the effects of electronic noise that is present at all frequencies.) Roughly, how long is needed to send one picture? Note that the center-to-center separation of

adjacent pulses must be at least $2 \Delta t$ (the total width of any one pulse).

- 6.30** • (a) By inspection of Fig. 6.19, deduce the fundamental frequency in the Fourier series of $F(t)$. (b) What would you suggest is the highest harmonic that has a substantial amplitude? What is its frequency?

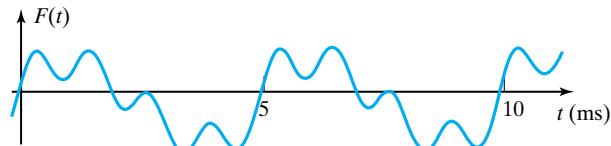


FIGURE 6.19

(Problem 6.30)

- 6.31** • Figure 6.20 shows a snapshot of a stretched garden hose that has been given an impulse at one end, so that a pulse is propagating along it with speed $v = 6 \text{ m/s}$. The width of the pulse is $\pm \Delta x$, with $\Delta x = 30 \text{ cm}$. (a) Find the approximate range Δk of wave numbers needed to build up this pulse. (b) When the pulse passes a certain point, it produces a brief vertical deflection $y = F(t)$. Find the width $\pm \Delta t$ of this pulse and the spread $\pm \Delta \omega$ when it is expressed as a superposition of sines and cosines of ωt .



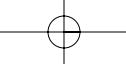
FIGURE 6.20

(Problem 6.31)

- 6.32** •• For a given periodic function $F(x)$, the coefficients A_n of its Fourier expansion can be found using the formulas (6.58) and (6.59) in Problem 6.35. [This is for the case of an even function, for which only cosine terms appear in the Fourier series. The general case involves sine terms as well, with coefficients given by (6.60) in Problem 6.36, but these do not appear in this problem.] Consider the periodic square pulse of Fig. 6.10 and verify that the Fourier coefficients are as claimed in (6.26) for $n \geq 1$ and that $A_0 = a/\lambda$. The height of the pulse is 1.

- 6.33** •• For theoretical purposes the best measure of the spread Δx of a pulse is the rms spread defined in (6.30), although this rms spread is sometimes rather less than one might guess. To illustrate this, think about the following: Consider a pulse whose probability density $P(x) = |\Psi(x, t)|^2$ (at one fixed time t) is as shown in Fig. 6.21(a). (The particle represented by this rectangular pulse is equally likely to be found anywhere between $x = -a$ and $x = a$. Since the total probability of finding the particle anywhere must be 1, the height of the pulse shown must be $1/2a$.) Since this pulse is centered at $x = 0$, the rms spread (6.30) is

$$\Delta x = \sqrt{\int_{-\infty}^{\infty} x^2 P(x) dx}$$



200 Chapter 6 • Matter Waves

Find Δx for the rectangular pulse.

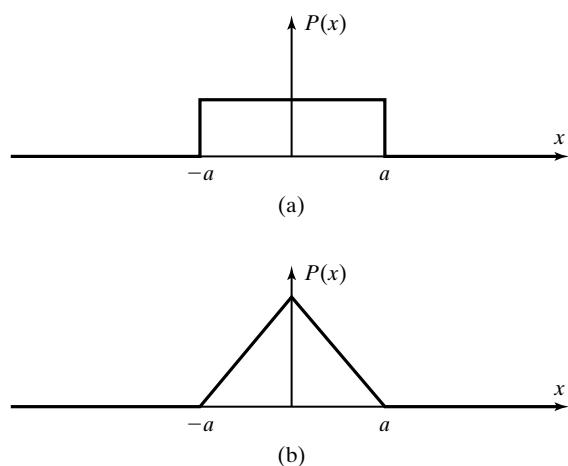


FIGURE 6.21

(Problems 6.33 and 6.34)

- 6.34** •• Do Problem 6.33, but for the triangular pulse of Fig. 6.21(b). The height of this pulse is $1/a$.

- 6.35** ••• The Fourier expansion theorem proves that any periodic function* $F(x)$ can be expanded in terms of sines and cosines. If the function happens to be even [$F(x) = F(-x)$], only cosines are needed and the expansion has the form

$$F(x) = \sum_{n=0}^{\infty} A_n \cos\left(\frac{2n\pi x}{\lambda}\right) \quad (6.57)$$

where λ is the period (or wavelength) of the function. In this problem you will see how to find the Fourier coefficients A_n . **(a)** Prove that

$$A_0 = \frac{1}{\lambda} \int_0^\lambda F(x) dx \quad (6.58)$$

[Hint: Integrate Eq. (6.57) from $x = 0$ to λ .]

(b) Prove that for $m > 0$,

$$A_m = \frac{2}{\lambda} \int_0^\lambda F(x) \cos\left(\frac{2m\pi x}{\lambda}\right) dx \quad (6.59)$$

where we have labeled the coefficient as A_m (rather than A_n) for reasons that will become apparent in your proof. [Hint: Multiply both sides of (6.57) by $\cos(2m\pi x/\lambda)$, and integrate from 0 to λ . Using the trig identities in Appendix B, you can prove that $\int_0^\lambda \cos(2m\pi x/\lambda) \cos(2m\pi x/\lambda) dx$ is zero if $m \neq n$ and equals $\lambda/2$ if $m = n$. In both parts of this problem you may assume that the integral of an infinite series, $\int [\sum g_n(x)] dx$, is the same as the series of integrals $\sum [\int g_n(x) dx]$.]

* $F(t)$ must satisfy some conditions of “reasonableness.” For example, the theorem is certainly true if $F(t)$ is continuous, although it is also true for many discontinuous functions as well.

- 6.36** ••• **(a)** Let $F(x)$ be a periodic function that is odd; that is, $F(x) = -F(-x)$. The Fourier expansion of such a function requires only sine functions:

$$F(x) = \sum_{n=0}^{\infty} B_n \sin\left(\frac{2n\pi x}{\lambda}\right)$$

Following the suggestions in Problem 6.35, prove that

$$B_m = \frac{2}{\lambda} \int_0^\lambda F(x) \sin\left(\frac{2m\pi x}{\lambda}\right) dx \quad (6.60)$$

(Note that the sine series has no $n = 0$ term, since $\sin 0 = 0$.) **(b)** Use this result to show that the Fourier coefficients B_n of the “sawtooth” function in Fig. 6.22 are zero for n even and that

$$B_n = (-1)^{(n-1)/2} \frac{8}{\pi^2 n^2} \quad \text{for } n \text{ odd}$$

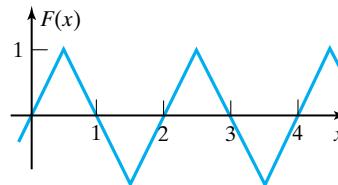


FIGURE 6.22

(Problems 6.36 and 6.53)

SECTION 6.8 (The Uncertainty Relation for Position and Momentum)

- 6.37** • A proton is known to be within an interval ± 6 fm (the radius of a large nucleus). Roughly, what is the minimum uncertainty in its velocity? Treat this problem as one-dimensional, and express your answer as a fraction of c .

- 6.38** • An air-rifle pellet has a mass of 20 g and a velocity of 100 m/s. If its velocity is known to an accuracy of $\pm 0.1\%$, what is the minimum possible uncertainty in the pellet’s position?

- 6.39** • The position of a 60-gram golf ball sitting on a tee is determined within $\pm 1 \mu\text{m}$. What is its minimum possible energy? Moving at the speed corresponding to this kinetic energy, how far would the ball move in a year?

- 6.40** • A classical physicist wants to use the Heisenberg microscope to disprove the uncertainty principle. To reduce the unknown momentum imparted to the electron, he reduces the lens diameter to one-third of its original value. How does this change Δp_x ? Δx ? Their product?

- 6.41** • Having failed to disprove the uncertainty principle in Problem 6.40, the physicist tries to reduce Δx by halving the object distance l . How does this change Δx ? Δp_x ? Their product?

- 6.42** •• Consider a proton confined to a region of typical nuclear dimensions, about 5 fm. **(a)** Use the uncertainty principle to estimate its minimum possible kinetic energy in MeV, assuming that it moves in only one dimension. **(b)** How would your result be modified

Problems for Chapter 6 **201**

if the proton were confined in a three-dimensional cube of side 5 fm? [See Eq. (6.40).] The actual kinetic energy of protons in nuclei is somewhat larger than this estimated minimum, being of order 10 MeV.

- 6.43** •• Consider an electron confined in a region of nuclear dimensions (about 5 fm). Find its minimum possible kinetic energy in MeV. Treat this problem as one-dimensional, and use the relativistic relation between E and p . (The large value you will find is a strong argument against the presence of electrons inside nuclei, since no known mechanism could contain an electron with this much energy.)

- 6.44** ••• The Heisenberg uncertainty principle applies to photons as well as to material particles. Thus, a photon confined to a very tiny box (size Δx) necessarily has a large uncertainty in momentum and energy (recall that for photons, $E = pc$) and hence a large average energy. Since energy is equivalent to mass, according to $E = mc^2$, a confined photon can create a large gravitational field. If Δx is sufficiently small, the energy density will be sufficiently large to create a black hole (Section 2.10). This size Δx is called the Planck length and defines the scale at which gravity and quantum mechanics are inextricably mixed. (The branch of physics that tries to combine quantum mechanics with gravity is called string theory.) **(a)** Show that the escape speed for a star of mass M and radius R is $v_{\text{esc}} = \sqrt{2GM/R}$. **(b)** By equating v_{esc} for a black hole to the speed of light, derive a formula for the Planck length in terms of c , G , and \hbar , and **(c)** show that its value is fantastically small, about 10^{-38} m. **(d)** What is the diameter of a proton in Planck lengths?

SECTION 6.9 (The Uncertainty Relation for Time and Energy)

- 6.45** • An excited state of a certain nucleus has a lifetime of 5×10^{-18} s. Find the minimum possible uncertainty in its energy.

- 6.46** • The subatomic particle called the $\Delta(1232)$ is an excited state of the proton, as we describe in Chapter 18. It decays into a pion and a proton with a mean life of order 10^{-23} s. What is the approximate uncertainty in its energy?

- 6.47** • The lifetime of the radioactive nucleus ^{235}U is about 1 billion years. Roughly, what is the uncertainty in its energy?

- 6.48** •• An unusually long-lived unstable atomic state has a lifetime of 1 ms. **(a)** Roughly, what is the minimum uncertainty in its energy? **(b)** Assuming that the photon emitted when this state decays is visible ($\lambda \approx 550$ nm), what are the uncertainty and fractional uncertainty in its wavelength?

SECTION 6.10 (Velocity of a Wave Packet)

- 6.49** • We have seen that for a nonrelativistic free particle, the particle's velocity v_{part} equals the velocity v_{pack} of

the corresponding wave packet [as given by (6.56)]. Prove that the same is true for a relativistic free particle. (Remember the Pythagorean relation between E and p .)

- 6.50** • **(a)** For waves on deep water (depth h much greater than wavelength λ) the wave velocity is given by* $v_{\text{wave}} = \sqrt{g/k}$. Prove that for these waves, the packet velocity v_{pack} is half the wave velocity v_{wave} . **(b)** In shallow water ($h \ll \lambda$), $v_{\text{wave}} = \sqrt{gh}$. Prove that, in this case, $v_{\text{pack}} = v_{\text{wave}}$.

- 6.51** • **(a)** Use Eqs. (6.54) and (6.55) to prove that

$$v_{\text{pack}} = v_{\text{wave}} - \lambda \frac{dv_{\text{wave}}}{d\lambda}$$

(b) When monochromatic light passes from a vacuum into glass, it is refracted, with blue light bending more than red. Use this observation to decide for which color v_{wave} is greater in glass — red or blue? **(c)** Use your results in parts (a) and (b) to prove that $v_{\text{pack}} < v_{\text{wave}}$ for light in glass.

- 6.52** • Prove that Eq. (6.51) follows from Eq. (6.50).

COMPUTER PROBLEMS

- 6.53** •• (Section 6.7) Consider the sawtooth wave shown in Fig. 6.22 with period 2 and maximum value 1. Because this function is odd, its Fourier series contains only sine functions:

$$F(x) = \sum_{n=1}^{\infty} B_n \sin(n\pi x)$$

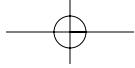
where the coefficients B_n are given in Problem 6.36(b). The function $F(x)$ can be approximated surprisingly well by using the sum of just a few terms in the Fourier series. To illustrate this, do the following: **(a)** Plot just the sum of the first two terms of the series for $0 \leq x \leq 4$. On the same picture, plot $F(x)$. **(b)** Repeat using the sum of the first five terms of the series ($n = 1, 3, 5, 7, 9$).

- 6.54** •• (Section 6.7) A certain function $F(x)$ has a Fourier cosine series like (6.25), with coefficients

$$A_n = \left(\frac{2}{n\pi}\right)^2 \left(1 - \cos\frac{n\pi}{2}\right)$$

for $n \geq 1$, and $A_0 = 1/4$. **(a)** Plot the sum of the first three terms ($n = 0, 1$, and 2) in the Fourier series for $-2 \leq x \leq 10$. **(b)** Do the same for the sum of the first six terms ($n = 0, 1, \dots, 5$), and **(c)** for the first eleven terms. **(d)** Can you describe the function $F(x)$?

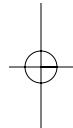
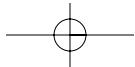
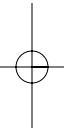
*Both of the formulas given here for v_{wave} ignore surface tension, which becomes important when λ is of order 1 cm or less.

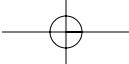


202 Chapter 6 • Matter Waves

6.55 •• (Section 6.7) Consider the periodic square pulse of Fig. 6.10, assuming that the pulse width $a = \frac{1}{3}$, the repeat length $\lambda = 1$, and the pulse height is 1. In the Fourier series (6.25) for this function, the coefficients A_n are given by (6.26) for $n > 0$, while $A_0 = a/\lambda$. **(a)** Using appropriate software, plot the sum of the

first two terms ($n = 0$ and 1) of the Fourier series for $0 \leq x \leq 2$. On the same picture, plot the square pulse itself. **(b)** Repeat, using the sum of the first 4 terms ($n = 0, 1, 2, 3$). **(c)** Repeat again, using the sum of the first 9 terms, and finally **(d)** for the sum of the first 31 terms. Do your pictures agree with Fig. 6.11?





Chapter 7

The Schrödinger Equation in One Dimension

- 7.1 Introduction
- 7.2 Classical Standing Waves
- 7.3 Standing Waves in Quantum Mechanics; Stationary States
- 7.4 The Particle in a Rigid Box
- 7.5 The Time-Independent Schrödinger Equation
- 7.6 The Rigid Box Again
- 7.7 The Free Particle
- 7.8 The Nonrigid Box
- 7.9 The Simple Harmonic Oscillator*
- 7.10 Tunneling*
- 7.11 The Time-Dependent Schrödinger Equation*

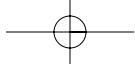
Problems for Chapter 7

* Sections marked with a star can be omitted without significant loss of continuity.

7.1 Introduction

In classical mechanics the state of motion of a particle is specified by giving the particle's position and velocity. In quantum mechanics the state of motion of a particle is specified by giving the wave function. In either case the fundamental question is to predict how the state of motion will evolve as time goes by, and in each case the answer is given by an *equation of motion*. The classical equation of motion is Newton's second law, $\mathbf{F} = m\mathbf{a}$; if we know the particle's position and velocity at time $t = 0$, Newton's second law determines the position and velocity at any other time. In quantum mechanics the equation of motion is the *time-dependent Schrödinger equation*. If we know a particle's wave function at $t = 0$, the time-dependent Schrödinger equation determines the wave function at any other time.

The time-dependent Schrödinger equation is a partial differential equation, a complete understanding of which requires more mathematical preparation than we are assuming here. Fortunately, the majority of interesting problems in quantum mechanics do not require use of the equation in its full generality. By far the most interesting states of any quantum system are those states in which the system has a definite total energy, and it turns out that for these states the wave function is a *standing wave*, analogous to the familiar standing waves on a string. When the time-dependent Schrödinger equation is applied to these standing waves, it reduces to a simpler equation called the *time-independent Schrödinger equation*. We will need only this



204 Chapter 7 • The Schrödinger Equation in One Dimension

time-independent equation, which will let us find the wave functions of the standing waves and the corresponding allowed energies. Because we will be using only the time-independent Schrödinger equation we will often refer to it as just “the Schrödinger equation.” Nevertheless, you should know that there are really two Schrödinger equations (the time-dependent and the time-independent). Unfortunately, it is almost universal to refer to either as “the Schrödinger equation” and to let the context decide which is being discussed. In this book, however, “the Schrödinger equation” will always mean the simpler time-independent equation.

In Section 7.2 we review some properties of classical standing waves, using waves on a uniform, stretched string as our example. In Section 7.3 we discuss quantum standing waves. Then, in Section 7.4, we show how the familiar properties of classical standing waves let one find the allowed energies of one simple quantum system, namely a particle that moves freely inside a perfectly rigid box.

Using our experience with the wave functions of a particle in a rigid box we next write down the time-independent Schrödinger equation, with which one can, in principle, find the allowed energies and wave functions for any system. Then in Sections 7.6 to 7.9 we use the Schrödinger equation to find the allowed energies of various simple systems.

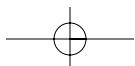
Sections 7.10 and 7.11 treat two further topics, quantum tunneling and the time-dependent Schrödinger equation. While both of these are very important, we will not be using these ideas until later (the former in Chapters 14 and 17, and the latter in Chapter 11), so you could skip these sections on your first reading.

Throughout this chapter we treat particles that move nonrelativistically in one dimension. All real systems are, of course, three-dimensional. Nevertheless, just as is the case in classical mechanics, it is a good idea to start with the simpler problem of a particle confined to move in just one dimension. In the classical case it is easy to find examples of systems that are at least approximately one-dimensional — a railroad car on a straight track, a bead threaded on a taut string. In quantum mechanics there are fewer examples of one-dimensional systems. However, we can for the moment imagine an electron moving along a very narrow wire.* The main importance of one-dimensional systems is that they provide a good introduction to three-dimensional systems, and that several one-dimensional solutions find direct application in three-dimensional problems.

7.2 Classical Standing Waves

We start with a review of some properties of classical standing waves in one dimension. We could discuss waves on a string, for which the wave function is the string’s transverse displacement $y(x, t)$; or we might consider sound waves, for which the wave function is the pressure variation, $p(x, t)$. If we considered electromagnetic waves, the wave function would be the electric field strength, $\mathcal{E}(x, t)$. In this section we choose to discuss waves on a string, but since our considerations apply equally to all waves, we will use the general notation $\Psi(x, t)$ to represent the wave function.

*More realistic examples include the motion of electrons along one axis in certain crystals and in some linear molecules.



Section 7.2 • Classical Standing Waves **205**

Let us consider first two sinusoidal traveling waves, one moving to the right,

$$\Psi_1(x, t) = B \sin(kx - \omega t)$$

(this is the wave sketched in Fig. 6.8) and the other moving to the left with the same amplitude,

$$\Psi_2(x, t) = B \sin(kx + \omega t)$$

The superposition principle guarantees that the sum of these two waves is itself a possible wave motion*:

$$\Psi(x, t) = \Psi_1(x, t) + \Psi_2(x, t) = B[\sin(kx - \omega t) + \sin(kx + \omega t)] \quad (7.1)$$

If we recall the important trigonometric identity (Appendix B)

$$\sin a + \sin b = 2 \sin \frac{a+b}{2} \cos \frac{a-b}{2}$$

we can rewrite the wave (7.1) as

$$\Psi(x, t) = 2B \sin kx \cos \omega t$$

or if we set $2B = A$,

$$\Psi(x, t) = A \sin kx \cos \omega t \quad (7.2)$$

A series of snapshots of the resultant wave (7.2) is sketched in Fig. 7.1. The important point to observe is that the resultant wave is not traveling to the right or left. At certain fixed points called **nodes**, where $\sin kx$ is zero, $\Psi(x, t)$ is always zero and the string is stationary. At any other point the string simply oscillates up and down in proportion to $\cos \omega t$, with amplitude $A \sin kx$. By superposing two traveling waves, we have formed a **standing wave**.

Because the string never moves at the nodes, we could clamp it at two nodes and remove the string outside the clamps, leaving a standing wave on a finite length of string as in Fig. 7.2. This is the kind of wave produced on a piano or guitar string when it sounds a pure musical tone.

If we now imagine a string clamped between two fixed points separated by a distance a , we can ask: What are the possible standing waves that can fit on the string? The answer is that a standing wave is possible, *provided* that it has nodes at the two fixed ends of the string. The distance between two adjacent nodes is $\lambda/2$, so the distance between any pair of nodes is an integer multiple of this, $n\lambda/2$. Therefore a standing wave fits on the string provided $n\lambda/2 = a$ for some integer n ; that is, if

$$\lambda = \frac{2a}{n}, \quad \text{where } n = 1, 2, 3, \dots \quad (7.3)$$

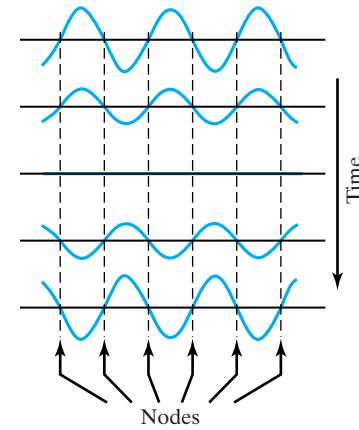


FIGURE 7.1

Five successive snapshots of the standing wave of Eq. (7.2). The nodes are points where the string remains stationary at all times. The distance between successive nodes is half a wavelength, $\lambda/2$.

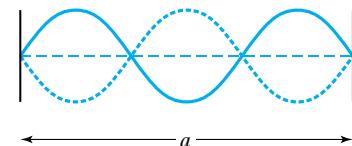
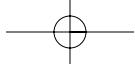


FIGURE 7.2

Three successive snapshots of a standing wave on a finite string, of length a , clamped at its two ends. The solid curve shows the string at maximum displacement; the dashed and dotted curves show it after successive quarter-cycle intervals.

*The superposition principle asserts that if Ψ_1 and Ψ_2 are possible waves, the same is true of $A\Psi_1 + B\Psi_2$ for any constants A and B . This important principle is true of any wave whose medium responds linearly to the disturbance. It applies to all the waves we will be considering.



206 Chapter 7 • The Schrödinger Equation in One Dimension

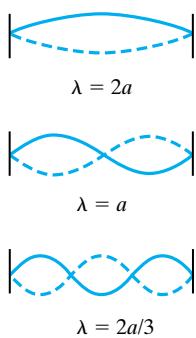


FIGURE 7.3

The first three possible standing waves on a string of length a , fixed at both ends. Each dashed curve is one half-cycle after the corresponding solid curve.

We see that the possible wavelengths of a standing wave on a string of length a are *quantized*, the allowed values being $2a$ divided by any positive integer. The first three of these allowed waves are sketched in Fig. 7.3.

It is important to recognize that the quantization of wavelengths arises from the requirement that the wave function must always be zero at the two fixed ends of the string. We refer to this kind of requirement as a **boundary condition**, since it relates to the boundaries of the system. We will find that for quantum waves, just as for classical waves, it is the boundary conditions that lead to quantization.

7.3 Standing Waves in Quantum Mechanics; Stationary States

Before we discuss quantum standing waves, we need to examine more closely the form of the classical standing wave (7.2):

$$\Psi(x, t) = A \sin kx \cos \omega t \quad (7.4)$$

This function is a *product* of one function of x (namely, $A \sin kx$) and one function of t (namely, $\cos \omega t$). We can emphasize this by rewriting (7.4) as

$$\Psi(x, t) = \psi(x) \cos \omega t \quad (7.5)$$

where we have used the capital letter Ψ for the full wave function $\Psi(x, t)$ and the lower case letter ψ for its **spatial part** $\psi(x)$. The spatial function $\psi(x)$ gives the full wave function $\Psi(x, t)$ at time $t = 0$ (since $\cos \omega t = 1$ when $t = 0$); more generally, at any time t the full wave function $\Psi(x, t)$ is $\psi(x)$ times the oscillatory factor $\cos \omega t$.

In our particular example (a wave on a uniform string) the spatial function $\psi(x)$ was a sine function

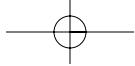
$$\psi(x) = A \sin kx \quad (7.6)$$

but in more general problems, such as waves on a nonuniform string, $\psi(x)$ can be a more complicated function of x . On the other hand, even in these more complicated problems the time dependence is still sinusoidal; that is, it is given by a sine or cosine function of t . The difference between the sine and the cosine is just a difference in the choice of origin of time. Thus either function is possible, and the general sinusoidal standing wave is a combination of both:

$$\Psi(x, t) = \psi(x)(a \cos \omega t + b \sin \omega t) \quad (7.7)$$

Different choices for the ratio of the coefficients a and b correspond to different choices of the origin of time. (See Problem 7.13.)

The standing waves of a quantum system have the same form (7.7), but with one important difference. For a classical wave, the function $\Psi(x, t)$ is, of course, a real number. (It would make no sense to say that the displacement of a string, or the pressure of a sound wave, had an imaginary part.) Therefore, the function $\psi(x)$ and the coefficients a and b in (7.7) are always real for any

Section 7.3 • Standing Waves in Quantum Mechanics; Stationary States **207**

classical wave.* In quantum mechanics, on the other hand, the wave function can be a complex number; and for quantum standing waves it usually is complex. Specifically, the time-dependent part of the wave function (7.7) always occurs in precisely the combination

$$\cos \omega t - i \sin \omega t \quad (7.8)$$

where i is the imaginary number $i = \sqrt{-1}$ (often denoted j by engineers). That is, the standing waves of a quantum particle have the form

$$\Psi(x, t) = \psi(x)(\cos \omega t - i \sin \omega t) \quad (7.9)$$

In Section 7.11 we will prove this from the time-dependent Schrödinger equation. For now, we simply assert that quantum standing waves have the sinusoidal time dependence of the particular combination of $\cos \omega t$ and $\sin \omega t$ in (7.9).

The form (7.9) can be simplified if we use Euler's formula from the theory of complex numbers (Problem 7.14),

$$\cos \theta + i \sin \theta = e^{i\theta} \quad (7.10)$$

This identity can be illustrated in the complex plane, as in Fig. 7.4, where the complex number $z = x + iy$ is represented by a point with coordinates x and y in the complex plane. Since the number $e^{i\theta}$ (with θ any real number) has coordinates $\cos \theta$ and $\sin \theta$, we see from Pythagoras' theorem that its absolute value is 1:

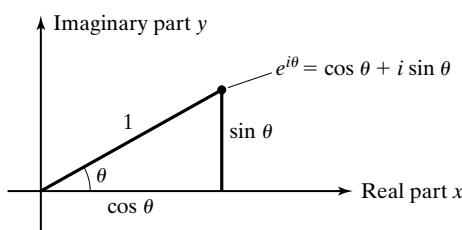
$$|e^{i\theta}| = \sqrt{(\cos \theta)^2 + (\sin \theta)^2} = 1$$

Thus the complex number $e^{i\theta}$ lies on a circle of radius 1, with polar angle θ as shown. Notice that since $\cos(-\theta) = \cos \theta$ and $\sin(-\theta) = -\sin \theta$,

$$\cos \theta - i \sin \theta = e^{-i\theta}$$

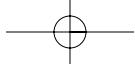
Returning to (7.9) and using the identity (7.10), we can write for the general standing wave of a quantum system

$$\Psi(x, t) = \psi(x)e^{-i\omega t} \quad (7.11)$$

**FIGURE 7.4**

The complex number $e^{i\theta} = \cos \theta + i \sin \theta$ is represented by a point with coordinates $(\cos \theta, \sin \theta)$ in the complex plane. The absolute value of any complex number $z = x + iy$ is defined as $|z| = \sqrt{x^2 + y^2}$. Since $\cos^2 \theta + \sin^2 \theta = 1$, it follows that $|e^{i\theta}| = 1$.

*As you may know, it is sometimes a mathematical convenience to introduce a certain complex wave function. Nonetheless, in classical physics the actual wave function is always the real part of this complex function.



208 Chapter 7 • The Schrödinger Equation in One Dimension

Since this function has a definite angular frequency, ω , any quantum system with this wave function has a definite energy given by the de Broglie relation $E = \hbar\omega$ (6.23). Conversely, any quantum system that has a definite energy has a wave function of the form (7.11) — a statement we will prove in Section 7.11.

We saw in Chapter 6 that the probability density associated with a quantum wave function $\Psi(x, t)$ is the absolute value squared, $|\Psi(x, t)|^2$. For the complex standing wave (7.11) this has a remarkable property:

$$|\Psi(x, t)|^2 = |\psi(x)|^2 |e^{-i\omega t}|^2$$

or, since $|e^{-i\omega t}| = 1$,

$$|\Psi(x, t)|^2 = |\psi(x)|^2 \quad (\text{for quantum standing waves}) \quad (7.12)$$

That is, for a quantum standing wave, the probability density is *independent of time*. This is possible because the time-dependent part of the wave function,

$$e^{-i\omega t} = \cos \omega t - i \sin \omega t$$

is complex, with two parts that oscillate 90° out of phase; when one is growing, the other is shrinking in such a way that the sum of their squares is constant. Thus for a quantum standing wave, the distribution of matter (of electrons in an atom, or nucleons in a nucleus, for example) is time independent or *stationary*. For this reason a quantum standing wave is often called a **stationary state**. The stationary states are the modern counterpart of Bohr's stationary orbits and are precisely the states of definite energy. Because their charge distribution is static, atoms in stationary states do not radiate.*

An important practical consequence of (7.12) is that in most problems the only interesting part of the wave function $\Psi(x, t)$ is its spatial part $\psi(x)$. We will see that a large part of quantum mechanics is devoted to finding the possible spatial functions $\psi(x)$ and their corresponding energies. Our principal tool in finding these will be the time-independent Schrödinger equation.

7.4 The Particle in a Rigid Box

Before we write down the Schrödinger equation, we consider a simple example that we can solve using just our experience with standing waves on a string. We consider a particle that is confined to some finite interval on the x -axis, and moves freely inside that interval — a situation we describe as a **one-dimensional rigid box** and (for reasons we explain later) is often called the infinite square well. For example, in classical mechanics we could consider a bead on a frictionless straight thread between two rigid knots; the bead can move freely between the knots, but cannot escape outside them. In quantum mechanics we can imagine an electron inside a length of very thin conducting wire; to a fair approximation, the electron would move freely back and forth inside the wire, but could not escape from it.

Let us consider, then, a quantum particle of mass m moving nonrelativistically in a one-dimensional rigid box of length a , with no forces acting on it between $x = 0$ and $x = a$. The absence of forces means that the potential

*Of course, atoms do radiate from excited states, but as we discuss in Chapter 11, this is always because some external influence disturbs the stationary state.

Section 7.4 • The Particle in a Rigid Box **209**

energy is constant inside the box, and we are free to choose that constant to be zero. Therefore, its total energy is just its kinetic energy. In quantum mechanics it is almost always more convenient to think of the kinetic energy as $p^2/2m$, rather than $\frac{1}{2}mv^2$, because of the de Broglie relation, $\lambda = h/p$ (6.1), between the momentum and wavelength. Therefore, we write the energy as

$$E = K = \frac{p^2}{2m} \quad (7.13)$$

As we have said, the states of definite energy are the standing waves. Therefore, to find the allowed energies, we must find the possible standing waves for the particle's wave function $\Psi(x, t)$. We have asserted that the standing waves have the form

$$\Psi(x, t) = \psi(x)e^{-i\omega t} \quad (7.14)$$

By analogy, with waves on a string, one might guess that the spatial function $\psi(x)$ will be a sinusoidal function inside the box; that is, $\psi(x)$ should have the form $\sin kx$ or $\cos kx$ or a combination of both:

$$\psi(x) = A \sin kx + B \cos kx \quad (7.15)$$

for $0 \leq x \leq a$. (We make no claim to have proved this; but it is certainly a reasonable guess, and we will prove it in Section 7.6.)

Since it is impossible for the particle to escape from the box, the wave function must be zero outside; that is, $\psi(x) = 0$ when $x < 0$ and when $x > a$. If we make the plausible (and, again, correct) assumption that $\psi(x)$ is continuous, then it must also vanish at $x = 0$ and $x = a$:

$$\psi(0) = \psi(a) = 0 \quad (7.16)$$

These are the boundary conditions that the wave function (7.15) must satisfy. Notice that these boundary conditions are identical to those for a classical wave on a string clamped at $x = 0$ and $x = a$.

From (7.15) we see that $\psi(0) = B$. Thus the wave function (7.15) can satisfy the boundary condition (7.16) only if the coefficient B is zero; that is, the condition $\psi(0) = 0$ restricts $\psi(x)$ to have the form

$$\psi(x) = A \sin kx \quad (7.17)$$

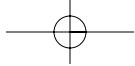
Next, the boundary condition that $\psi(a) = 0$ requires that

$$A \sin ka = 0 \quad (7.18)$$

which implies that*

$$ka = \pi, \text{ or } 2\pi, \text{ or } 3\pi, \dots \quad (7.19)$$

*Strictly speaking, (7.18) implies either that k satisfies (7.19) or that $A = 0$; but if $A = 0$, then $\psi = 0$ for all x , and we get no wave at all. Thus, only the solution (7.19) corresponds to a particle in a box. Notice also that there is no reason to include negative integer values of n in (7.20) since $\sin(-kx)$ is just a multiple of $\sin(kx)$.



210 Chapter 7 • The Schrödinger Equation in One Dimension

or

$$k = \frac{n\pi}{a}, \quad n = 1, 2, 3, \dots \quad (7.20)$$

We conclude that the only standing waves that satisfy the boundary conditions (7.16) have the form $\psi(x) = A \sin kx$ with k given by (7.20). In terms of wavelength, this condition implies that

$$\lambda = \frac{2\pi}{k} = \frac{2a}{n} \quad n = 1, 2, 3, \dots \quad (7.21)$$

which is precisely the condition (7.3) for standing waves on a string. This is, of course, not an accident. In both cases, the quantization of wavelengths arose from the boundary condition that the wave function must be zero at $x = 0$ and $x = a$.

For our present discussion, the important point is that quantization of wavelength λ implies quantization of momentum, and hence also of energy. Specifically, substituting (7.21) into the de Broglie relation $p = h/\lambda$, we find that

$$p = \frac{nh}{2a} = \frac{n\pi\hbar}{a} \quad n = 1, 2, 3, \dots \quad (7.22)$$

Since $E = K + U$ and $U = 0$ in this case, we have $E = p^2/2m$. Therefore, (7.22) means that the allowed energies for a particle in a one-dimensional rigid box are

$$E_n = \frac{n^2\pi^2\hbar^2}{2ma^2} \quad n = 1, 2, 3, \dots \quad (7.23)$$

The lowest energy for our particle, termed the **ground-state energy**, is obtained when $n = 1$ and is

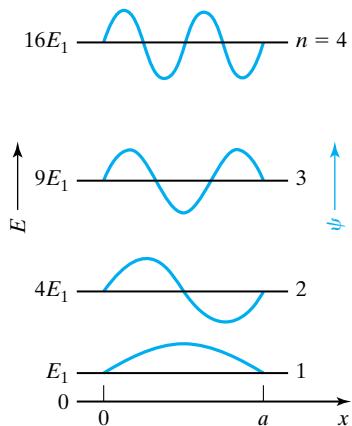


FIGURE 7.5

A composite picture showing the first four energy levels and wave functions for a particle in a rigid box. Each horizontal line indicates an energy level and is also used as the axis for a plot of the corresponding wave function.

$$E_1 = \frac{\pi^2\hbar^2}{2ma^2} \quad (7.24)$$

This is consistent with the lower bound derived from the Heisenberg uncertainty principle in Chapter 6, where we argued — see (6.39) — that for a particle confined in a region of length a ,

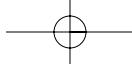
$$E \geq \frac{\hbar^2}{2ma^2} \quad (7.25)$$

For our particle in a rigid box, the actual minimum energy (7.24) is larger than the lower bound (7.25) by a factor of $\pi^2 \approx 10$.

In terms of the ground-state energy E_1 , the energy of the n th level (7.23) is

$$E_n = n^2E_1 \quad n = 1, 2, 3, \dots \quad (7.26)$$

These energy levels are sketched in Fig. 7.5. Notice that (quite unlike those of the hydrogen atom) the energy levels are farther and farther apart as n increases and that E_n increases without limit as $n \rightarrow \infty$. The corresponding wave functions $\psi(x)$ (which look exactly like the standing waves on a string)

Section 7.5 • The Time-Independent Schrödinger Equation **211**

have been superimposed on the same picture, the wave function for each level being plotted on the line that represents its energy. Notice how the number of nodes of the wave functions increases steadily with energy; this is what one should expect since more nodes mean shorter wavelength (larger curvature of ψ) and hence larger momentum and kinetic energy.

The complete wave function $\Psi(x, t)$ for any of our standing waves has the form

$$\Psi(x, t) = \psi(x)e^{-i\omega t} = A \sin(kx)e^{-i\omega t}$$

We can rewrite this, using the identity (Problem 7.16)

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \quad (7.27)$$

to give

$$\Psi(x, t) = \frac{A}{2i}(e^{i(kx-\omega t)} - e^{-i(kx+\omega t)}) \quad (7.28)$$

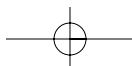
We see that our quantum standing wave (just like the classical standing wave of Section 7.2) can be expressed as the sum of two traveling waves, one moving to the right and one to the left. The wave moving to the right represents a particle with momentum $\hbar k$ *directed to the right*, and that moving to the left, a particle with momentum of the same magnitude $\hbar k$ *directed to the left*. Thus a particle in one of our stationary states has a definite magnitude, $\hbar k$, for its momentum but is an equal superposition of momenta in either direction. This corresponds to the result that *on average* a classical particle is equally likely to be moving in either direction as it bounces back and forth inside a rigid box.

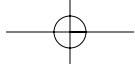
7.5 The Time-Independent Schrödinger Equation

Our discussion of the particle in a rigid box depended on some guessing as to the form of the spatial wave function $\psi(x)$. There are very few problems where this kind of guesswork is possible, and no problems where it is entirely satisfying. What we need is the equation that determines $\psi(x)$ in *any* problem, and this equation is the time-independent Schrödinger equation. Like all basic laws of physics, the Schrödinger equation cannot be *derived*. It is simply a relation, like Newton's second law, that experience has shown to be true. Thus a legitimate procedure would be simply to state the equation and to start using it. Nevertheless, it may be helpful to offer some arguments that *suggest* the equation, and this is what we will try to do.

Almost all laws of physics can be expressed as *differential equations*, that is, as equations that involve the variable of interest and some of its derivatives. The most familiar example is Newton's second law for a single particle, which we can write as

$$m \frac{d^2x}{dt^2} = \sum F \quad (7.29)$$





212 Chapter 7 • The Schrödinger Equation in One Dimension

If, for example, the particle in question were immersed in a viscous fluid that exerted a drag force $-bv$, and attached to a spring that exerted a restoring force $-kx$, then (7.29) would read

$$m \frac{d^2x}{dt^2} = -b \frac{dx}{dt} - kx \quad (7.30)$$

This is a differential equation for the particle's position x as a function of time t , and since the highest derivative involved is the second derivative, the equation is called a *second-order differential equation*.

The equation of motion for classical waves (which is often not discussed in an introductory physics course) is a differential equation. It is therefore natural to expect the equation that determines the possible standing waves of a quantum system to be a differential equation. Since we already know the form of the wave functions for a particle in a rigid box, what we will do is examine these wave functions and try to spot a simple differential equation that they satisfy and that we can generalize to more complicated systems.

We saw in Section 7.4 that the spatial wave functions for a particle in a rigid box have the form

$$\psi(x) = A \sin kx \quad (7.31)$$

To find a differential equation that this function satisfies, we naturally differentiate it, to give

$$\frac{d\psi}{dx} = kA \cos kx \quad (7.32)$$

There are several ways in which we could relate $\cos kx$ in (7.32) to $\sin kx$ in (7.31) and hence obtain an equation connecting $d\psi/dx$ with ψ . However, a simpler course is to differentiate a second time to give

$$\frac{d^2\psi}{dx^2} = -k^2 A \sin kx \quad (7.33)$$

Comparing (7.33) and (7.31), we see at once that $d^2\psi/dx^2$ is proportional to ψ ; specifically,

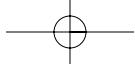
$$\frac{d^2\psi}{dx^2} = -k^2 \psi \quad (7.34)$$

We can rewrite k^2 in (7.34) in terms of the particle's kinetic energy, K . We know that $p = \hbar k$. Therefore,

$$K = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \quad (7.35)$$

hence

$$k^2 = \frac{2mK}{\hbar^2} \quad (7.36)$$

Section 7.5 • The Time-Independent Schrödinger Equation **213**

Thus, we can write (7.34) as

$$\frac{d^2\psi}{dx^2} = -\frac{2mK}{\hbar^2}\psi \quad (7.37)$$

which gives us a second-order differential equation satisfied by the wave function $\psi(x)$ of a particle in a rigid box.

The particle in a rigid box is an especially simple system, with potential energy equal to zero throughout the region where the particle moves. It is not at all obvious how the equation (7.37) should be generalized to include the possibility of a nonzero potential energy, $U(x)$, which may vary from point to point. However, since the kinetic energy K is the difference between the total energy E and the potential energy $U(x)$, it is perhaps natural to replace K in (7.37) by

$$K = E - U(x) \quad (7.38)$$

This gives us the differential equation*

$$\frac{d^2\psi}{dx^2} = \frac{2m}{\hbar^2} [U(x) - E]\psi \quad (7.39)$$

This differential equation is called the **Schrödinger equation** (**time-independent Schrödinger equation**, in full), to honor the Austrian physicist, Erwin Schrödinger, who first published it in 1926. Like us, Schrödinger had no way to *prove* that his equation was correct. All he could do was argue that the equation seemed reasonable and that its predictions should be tested against experiment. In the 80 years or so since then, it has passed this test repeatedly. In particular, Schrödinger himself showed that it predicts correctly the energy levels of the hydrogen atom, as we describe in Chapter 8. Today, it is generally accepted that the Schrödinger equation is the correct basis of nonrelativistic quantum mechanics, in just the same way that Newton's second law is accepted as the basis of nonrelativistic classical mechanics.

The Schrödinger equation as written in (7.39) applies to one particle moving in one dimension. We will need to generalize it later to cover systems of several particles, in two or three dimensions. Nevertheless, the general procedure for using the equation is the same in all cases. Given a system whose stationary states and energies we want to know, we must first find the potential energy function $U(x)$. For example, a particle held in equilibrium at $x = 0$ by a force obeying Hooke's law ($F = -kx$) has potential energy

$$U(x) = \frac{1}{2}kx^2 \quad (7.40)$$

Erwin Schrödinger
(1887–1961, Austrian)

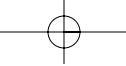


After learning of de Broglie's matter waves, Schrödinger proposed the equation — the Schrödinger equation — that governs the waves' behavior and earned him the 1933 Nobel Prize in physics. He left Austria after Hitler's invasion and became a professor in Dublin, Ireland. A person with remarkably broad interests, he was an ardent student of Italian painting and botany, as well as chemistry and physics. Late in his career, he became a pioneer in the new field of biophysics and wrote a popular book entitled *What is Life?*.

*In more advanced texts this equation is usually written in the form

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + U(x)\psi = E\psi$$

Because the differential operator $-(\hbar^2/2m)d^2/dx^2$ is intimately connected with the kinetic energy, this way of writing the **Schrödinger** equation is perhaps easier to remember because it looks like $K + U = E$. Nevertheless, for the applications in this book, the form (7.39) is the most convenient, and we will almost always write it this way.



214 Chapter 7 • The Schrödinger Equation in One Dimension

An electron in a hydrogen atom has

$$U(r) = -\frac{ke^2}{r} \quad (7.41)$$

(We will return to this three-dimensional example in Chapter 8.) Once we have identified $U(x)$, the Schrödinger equation (7.39) becomes a well-defined equation that we can try to solve.* In most cases, it turns out that for many values of the energy E the Schrödinger equation *has no solutions* (no acceptable solutions, satisfying the particular conditions of the problem, that is). This is exactly what leads to the quantization of energies. Those values of E for which the Schrödinger equation has no solution are not allowed energies of the system. Conversely, those values of E for which there is a solution *are* allowed energies, and the corresponding solutions $\psi(x)$ give the spatial wave functions of these stationary states.

As we hinted in the last paragraph, there are usually certain conditions that the wave function $\psi(x)$ must satisfy to be an acceptable solution of the Schrödinger equation. First, there may be boundary conditions on $\psi(x)$, for example, the condition that $\psi(x)$ must vanish at the walls of a perfectly rigid box. In addition, there are certain general restrictions on $\psi(x)$; for example, as we anticipated in Section 7.4, $\psi(x)$ must always be *continuous*, and in most problems its first derivative must also be continuous. When we speak of an acceptable solution of the Schrödinger equation, we mean a solution that satisfies all the conditions appropriate to the problem at hand.

In this section you may have noticed that in quantum mechanics it is the potential energy $U(x)$ that appears in the basic equation, whereas in classical mechanics it is the *force* F . Of course, U and F are closely related: F being the derivative of U , U being the integral of F . Nevertheless, it is an important difference of emphasis that quantum mechanics focuses primarily on potential energies, whereas Newtonian mechanics focuses on forces.

7.6 The Rigid Box Again

As a first application of the Schrödinger equation, we use it to rederive the allowed energies of a particle in a rigid box and check that we get the same answers as before.[†] The first step in applying the Schrödinger equation to any system is to identify the potential-energy function $U(x)$. Inside the box we can choose the potential energy to be zero, and outside the box it is infinite. This is the mathematical expression of our idealized *perfectly* rigid box — no finite amount of energy can remove the particle from it. Thus

$$U(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq a \\ \infty & \text{for } x < 0 \text{ and } x > a \end{cases} \quad (7.42)$$

*The necessity of identifying U before one can solve the Schrödinger equation corresponds to the necessity of identifying the total force F on a classical particle before one can solve Newton's second law, $F = ma$. The two are, of course, closely related since $F = -dU/dx$.

[†] You may reasonably object that it is circular to apply the Schrödinger equation to a particle in a box, when we used the latter to derive the former. Nevertheless, it is a legitimate consistency check, as well as an instructive exercise, to see how the Schrödinger equation gives back the known energies and wave functions.

Section 7.6 • The Rigid Box Again **215**

That $U(x) = \infty$ outside the box implies that the particle can never be found there and hence that the wave function $\psi(x)$ must be zero when $x < 0$ and when $x > a$. The potential-energy function is described as an infinitely deep potential well or an infinite square well because of the square (90°) angles at the bottom of the well. The continuity of $\psi(x)$ then requires that

$$\psi(0) = \psi(a) = 0 \quad (7.43)$$

(all of which we had argued in Section 7.4). Inside the box, where $U(x) = 0$, the Schrödinger equation (7.39) reduces to

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2}\psi \quad \text{for } 0 \leq x \leq a \quad (7.44)$$

This is the differential equation whose solutions we must investigate. In particular, we want to find those values of E for which it has a solution satisfying the boundary conditions (7.43).

Before solving (7.44), we remark that it is a nuisance, both for the printer of a book and for the student taking notes, to keep writing the symbols $d\psi/dx$ and $d^2\psi/dx^2$. For this reason, we introduce the shorthand

$$\psi' \equiv \frac{d\psi}{dx} \quad \text{and} \quad \psi'' \equiv \frac{d^2\psi}{dx^2}$$

From now on we will use this notation whenever convenient. In particular, we rewrite (7.44) as

$$\psi''(x) = -\frac{2mE}{\hbar^2}\psi(x) \quad (7.45)$$

We now consider whether there is an acceptable solution of (7.45) for any particular value of E , starting with the case that E is negative. (We do not expect any states with $E < 0$, since then E would be less than the minimum potential energy. But we have already encountered several unexpected consequences of quantum mechanics, and we should check this possibility.) If E were negative, the coefficient $-2mE/\hbar^2$ on the right of (7.45) would be *positive* and we could call it α^2 , where

$$\alpha = \frac{\sqrt{-2mE}}{\hbar} \quad (7.46)$$

With this notation, (7.45) becomes

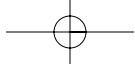
$$\psi''(x) = \alpha^2\psi(x) \quad (7.47)$$

The simplification of rewriting (7.45) in the form (7.47) has the disadvantage of requiring a new symbol (namely α); but it has the important advantage of letting us focus on the mathematical structure of the equation.

Equation (7.47) is a second-order differential equation, which has the solutions (Problem 7.22) $e^{\alpha x}$ and $e^{-\alpha x}$ or any combination of these,

$$\psi(x) = Ae^{\alpha x} + Be^{-\alpha x} \quad (7.48)$$

where A and B are any constants, real or complex.



216 Chapter 7 • The Schrödinger Equation in One Dimension

It is important in what follows that (7.48) is the most general solution of (7.47), that is, that *every* solution of (7.47) has the form (7.48). This follows from a theorem about second-order differential equations of the same type as the one-dimensional Schrödinger equation.* This theorem states three facts: First, these equations always have two independent solutions. For example, $e^{\alpha x}$ and $e^{-\alpha x}$ are two independent solutions of (7.47).† Second, if $\psi_1(x)$ and $\psi_2(x)$ denote two such independent solutions, then the linear combination

$$A\psi_1(x) + B\psi_2(x) \quad (7.49)$$

is also a solution, for any constants A and B . (This is the superposition principle.) Third, given two independent solutions $\psi_1(x)$ and $\psi_2(x)$, *every* solution can be expressed as a linear combination of the form (7.49). These three properties are illustrated in Problems 7.22 to 7.28.

That the general solution of a second-order differential equation contains two arbitrary constants is easy to understand: A second-order differential equation amounts to a statement about the second derivative ψ'' ; to find ψ , one must somehow accomplish two integrations, which should introduce two constants of integration; and this is what the two arbitrary constants A and B in (7.49) are. The theorem above is very useful in seeking solutions of such differential equations. If, by any means, we can spot two independent solutions, we are assured that *every* solution is a combination of these two. Since $e^{\alpha x}$ and $e^{-\alpha x}$ are independent solutions of (7.47), it follows from the theorem that the most general solution is (7.48).

Equation (7.48) gives all solutions of the Schrödinger equation for negative values of E . The important question now is whether any of these solutions could satisfy the required boundary conditions (7.43), and the answer is “no.” With $\psi(x)$ given by (7.48), the condition that $\psi(0) = 0$ implies that

$$A + B = 0$$

while the requirement that $\psi(a) = 0$ implies that

$$Ae^{aa} + Be^{-aa} = 0$$

One can verify (Problem 7.23) that the only values of A and B that satisfy these two simultaneous equations are $A = B = 0$. That is, if $E < 0$, the only solution of the Schrödinger equation that satisfies the boundary conditions is the zero function. In other words, with $E < 0$, there can be no standing waves, so negative values of E are not allowed. A similar argument gives the same conclusion for $E = 0$.

Let us next see if the Schrödinger equation (7.45) has any acceptable solutions for positive energies (as we expect it does). With $E > 0$, the coefficient $-2mE/\hbar^2$ on the right of (7.45) is negative and can conveniently be called $-k^2$ where

$$k = \frac{\sqrt{2mE}}{\hbar} \quad (7.50)$$

*To be precise, ordinary second-order differential equations that are linear and homogeneous.

†When we say that two functions are *independent*, we mean that neither function is just a constant multiple of the other. For example, $e^{\alpha x}$ and $e^{-\alpha x}$ are independent, but $e^{\alpha x}$ and $2e^{\alpha x}$ are not; similarly $\sin x$ and $\cos x$ are independent, but $5 \cos x$ and $\cos x$ are not.

With this notation, the Schrödinger equation reads

$$\psi''(x) = -k^2\psi(x) \quad (7.51)$$

This differential equation has the solutions $\sin kx$ and $\cos kx$, or any combination of both:

$$\psi(x) = A \sin kx + B \cos kx \quad (7.52)$$

(see Example 7.1 below). This is exactly the form of the wave function that we assumed at the beginning of Section 7.4. The important difference is that in Section 7.4 we could only guess the form (7.52), whereas we have now *derived* it from the Schrödinger equation. From here on, the argument follows precisely the argument given before. As we saw, the boundary condition $\psi(0) = 0$ requires that the coefficient B in (7.52) be zero, whereas the condition that $\psi(a) = 0$ can be satisfied without A being zero, provided that ka is an integer multiple of π (so that $\sin ka = 0$); that is,

$$k = \frac{n\pi}{a}$$

or, from (7.50),

$$E = \frac{\hbar^2 k^2}{2m} = n^2 \frac{\pi^2 \hbar^2}{2ma^2}$$

exactly as before.

Example 7.1

Verify explicitly that the function (7.52) is a solution of the Schrödinger equation (7.51) for any values of the constants A and B . [This illustrates part of the theorem stated in connection with (7.49).]

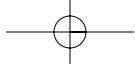
To verify that a given function satisfies an equation, one must substitute the function into one side of the equation and then manipulate it until one arrives at the other side. Thus, for the proposed solution (7.52),

$$\begin{aligned} \psi''(x) &= \frac{d^2}{dx^2}(A \sin kx + B \cos kx) \\ &= \frac{d}{dx}(kA \cos kx - kB \sin kx) \\ &= -k^2 A \sin kx - k^2 B \cos kx \\ &= -k^2(A \sin kx + B \cos kx) \\ &= -k^2\psi(x) \end{aligned}$$

and we conclude that the proposed solution does satisfy the desired equation.

There is one loose end in our discussion of the particle in a rigid box that we can now dispose of. We have seen that the stationary states have wave functions

$$\psi(x) = A \sin \frac{n\pi x}{a} \quad (7.53)$$



218 Chapter 7 • The Schrödinger Equation in One Dimension

but we have not yet found the constant A . Whatever the value of A , the function (7.53) satisfies the Schrödinger equation and the boundary conditions. Clearly, therefore, neither the Schrödinger equation nor the boundary conditions fix the value of A .

To see what does fix A , recall that $|\psi(x)|^2$ is the probability density for finding the particle at x . This means, in the case of a one-dimensional system, that $|\psi(x)|^2 dx$ is the probability P of finding the particle between x and $x + dx$.

$$P(\text{between } x \text{ and } x + dx) = |\psi(x)|^2 dx \quad (7.54)$$

Since the total probability of finding the particle *anywhere* must be 1, it follows that

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1 \quad (7.55)$$

This relation is called the **normalization condition** and a wave function that satisfies it is said to be **normalized**. It is the condition (7.55) that fixes the value of the constant A , which is therefore called the **normalization constant**.

In the case of the rigid box, $\psi(x)$ is zero outside the box; therefore, (7.55) can be rewritten as

$$\int_0^a |\psi(x)|^2 dx = 1 \quad (7.56)$$

or, with the explicit form (7.53) for $\psi(x)$,

$$A^2 \int_0^a \sin^2\left(\frac{n\pi x}{a}\right) dx = 1 \quad (7.57)$$

The integral here turns out to be $a/2$ (Problem 7.29). Therefore, (7.57) implies that

$$\frac{A^2 a}{2} = 1 \quad (7.58)$$

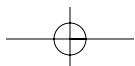
and hence that*

$$A = \sqrt{\frac{2}{a}} \quad (7.59)$$

We conclude that the normalized wave functions for the particle in a rigid box are given by

$$\psi(x) = \sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \quad (7.60)$$

* You may have noticed that, strictly speaking, the argument leading from (7.55) to (7.59) implies only that the *absolute value* of A is $\sqrt{2/a}$. However, since the probability density depends only on the absolute value of ψ , we are free to choose any value of A satisfying $|A| = \sqrt{2/a}$ (for example, $A = -\sqrt{2/a}$ or $i\sqrt{2/a}$); the choice (7.59) is convenient and customary.



Example 7.2

Consider a particle in the ground state of a rigid box of length a . **(a)** Find the probability density $|\psi|^2$. **(b)** Where is the particle most likely to be found? **(c)** What is the probability of finding the particle in the interval between $x = 0.50a$ and $x = 0.51a$? **(d)** What is it for the interval $[0.75a, 0.76a]$? **(e)** What would be the average result if the position of a particle in the ground state were measured many times?

- (a)** The probability density is just $|\psi(x)|^2$, where $\psi(x)$ is given by (7.60) with $n = 1$. Therefore, it is

$$|\psi(x)|^2 = \frac{2}{a} \sin^2\left(\frac{\pi x}{a}\right) \quad (7.61)$$

which is sketched in Fig. 7.6.

- (b)** The most probable value x , is the value of x for which $|\psi(x)|^2$ is maximum. From Fig. 7.6 this is clearly seen to be

$$x_{\text{mp}} = a/2 \quad (7.62)$$

- (c)** The probability of finding the particle in any small interval from x to $x + \Delta x$ is given by (7.54) as

$$P(\text{between } x \text{ and } x + \Delta x) \approx |\psi(x)|^2 \Delta x \quad (7.63)$$

(This is exact in the limit $\Delta x \rightarrow 0$ and is therefore a good approximation for any small interval Δx .) Thus, the two probabilities are

$$\begin{aligned} P(0.50a \leq x \leq 0.51a) &\approx |\psi(0.50a)|^2 \Delta x = \frac{2}{a} \sin^2\left(\frac{\pi}{2}\right) \times 0.01a \\ &= 0.02 = 2\% \end{aligned}$$

- (d)** and, similarly,

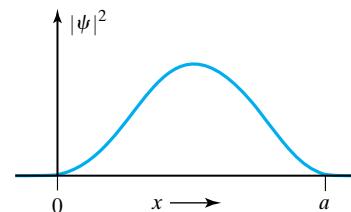
$$P(0.75a \leq x \leq 0.76a) \approx \frac{2}{a} \sin^2\left(\frac{3\pi}{4}\right) \times 0.01a = 0.01 = 1\%$$

- (e)** The average result if we measure the position many times (always with the particle in the same state) is the integral, over all possible positions, of x times the probability of finding the particle at x :

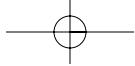
$$\langle x \rangle = \int_0^a x |\psi(x)|^2 dx \quad (7.64)$$

(If you are not familiar with this argument, see the following paragraphs.) This average value $\langle x \rangle$ (also denoted \bar{x} or x_{av}) is often called the **expectation value** of x . (But note that it is not the value we expect in any one measurement; it is rather the average value expected after many measurements.) In the present case

$$\langle x \rangle = \frac{2}{a} \int_0^a x \sin^2\left(\frac{\pi x}{a}\right) dx \quad (7.65)$$

**FIGURE 7.6**

The probability density $|\psi(x)|^2$ for a particle in the ground state of a rigid box. Inside the box, $|\psi|^2$ is given by (7.61); outside, it is zero.



220 Chapter 7 • The Schrödinger Equation in One Dimension

This integral can be evaluated to give (Problem 7.33)

$$\langle x \rangle = \frac{a}{2} \quad (7.66)$$

an answer that is easily understood from Fig. 7.6: Since $|\psi(x)|^2$ is symmetric about the middle position $x = a/2$, the average value must be $a/2$. We see from (7.62) and (7.66) that for the ground state of a rigid box, the most probable position x_{mp} and the mean position $\langle x \rangle$ are the same. We will see in the next example that x_{mp} and $\langle x \rangle$ are not always equal.

Expectation Values

In Example 7.2 we introduced the notion of the expectation value $\langle x \rangle$ of x . This is *not* the value of x expected in any one measurement; rather it is the average value expected if we repeat the measurement many times (always with the system in the same state). This kind of average comes up in many other branches of physics, especially in statistical mechanics, and is worth discussing in a little more detail. In particular, we want to justify the expression (7.64).

Suppose that we are interested in a quantity x that can take on various values with definite probabilities. The quantity x could be a continuous variable, such as the position of a quantum particle, or a discrete variable, such as the number of offspring of a female fruit fly chosen at random in a large colony of fruit flies. Let us consider first the discrete case: Suppose that the possible results of the measurement are $x_1, x_2, \dots, x_i, \dots$ and that these results occur with probabilities $P_1, P_2, \dots, P_i, \dots$. This statement means that if a large number N of statistically independent measurements are made, the number of measurements resulting in value x_i will be $n_i = P_i \cdot N$; in other words, $P_i = n_i/N$ is the fraction of the measurements that yield the value x_i . The average value of x is the sum of all the results of all the measurements divided by the total number N . Since n_i of the measurements produce the value x_i , this sum of all the measurement results is $\sum_i n_i x_i$, and the average value is

$$\langle x \rangle = \frac{1}{N} \sum_i n_i x_i$$

This expression can be rewritten in terms of the probabilities P_i as

$$\langle x \rangle = \frac{1}{N} \sum_i n_i x_i = \sum_i \frac{n_i}{N} x_i = \sum_i P_i x_i \quad (7.67)$$

If x is a continuous variable, the probability P_i is replaced with a probability increment $p(x) dx$, where $p(x)$ is the probability density. [For example, in the case of interest to us now, x is the position of a quantum particle and the probability density is $p(x) = |\psi(x)|^2$.] The sum (7.67) becomes an integral,

$$\langle x \rangle = \sum_i x_i P_i \rightarrow \langle x \rangle = \int x p(x) dx. \quad (7.68)$$

In particular, for a quantum particle $p(x) = |\psi(x)|^2$, and we have the expression (7.64) for the expectation value of the position. More generally, if we

measure x^2 or x^3 or any function $f(x)$, we can repeat the same argument, simply replacing x with $f(x)$, and conclude that

$$\langle f(x) \rangle = \int f(x)p(x) dx \quad (7.69)$$

Example 7.3

Answer the same questions as in Example 7.2 but for the first excited state of the rigid box.

The wave function is given by (7.60) with $n = 2$, so

$$|\psi(x)|^2 = \frac{2}{a} \sin^2\left(\frac{2\pi x}{a}\right)$$

This is plotted in Fig. 7.7, where it is clear that $|\psi(x)|^2$ has two equal maxima at

$$x_{mp} = \frac{a}{4} \quad \text{and} \quad \frac{3a}{4}$$

The expectation value $\langle x \rangle$ is easily found without actually doing any integration. Since $|\psi(x)|^2$ is symmetric about $x = a/2$, contributions to the integral (7.64) from either side of $x = a/2$ exactly balance one another, and we find the same answer as for the ground state

$$\langle x \rangle = \frac{a}{2}$$

The probabilities of finding the particle in any small intervals are given by (7.63) as

$$P(0.50a \leq x \leq 0.51a) \approx |\psi(0.50a)|^2 \Delta x = 0 \quad (7.70)$$

since* $\psi(0.50a) = 0$; and

$$P(0.75a \leq x \leq 0.76a) \approx |\psi(0.75a)|^2 \Delta x = \frac{2}{a} \sin^2\left(\frac{3\pi}{2}\right) \times 0.01a = 0.02$$

In particular, notice that although $x = a/2$ is the average value of x , the probability of finding the particle in the immediate neighborhood of $x = a/2$ is zero. This result, although a little surprising at first, is easily understood by reference to Fig. 7.7.

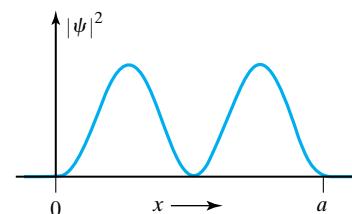
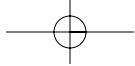


FIGURE 7.7

The probability density $|\psi(x)|^2$ for a particle in the first excited state ($n = 2$) of a rigid box.

*Note that the probability for the interval $[0.50a, 0.51a]$ is not exactly zero since the probability density $|\psi(x)|^2$ is zero only at the one point $0.50a$. The significance of (7.70) is really that the probability for this interval is *very small* compared to the probability for intervals of the same width elsewhere.



222 Chapter 7 • The Schrödinger Equation in One Dimension

7.7 The Free Particle

As a second application of the Schrödinger equation, we investigate the possible energies of a free particle; that is, a particle subject to no forces and completely unconfined (still in one dimension, of course). The potential energy of a free particle is constant and can be chosen to be zero. With this choice, we will show that the energy of the particle can have any positive value, $E \geq 0$. That is, the energy of a free particle is not quantized, and its allowed values are the same as those of a classical free particle.

To prove these assertions, we must write down the Schrödinger equation and find those E for which it has acceptable solutions. With $U(x) = 0$, the Schrödinger equation is

$$\psi''(x) = -\left(\frac{2mE}{\hbar^2}\right)\psi(x) \quad (7.71)$$

This is the same equation that we solved for a particle in a rigid box. However, there is an important difference since the free particle can be anywhere in the range

$$-\infty < x < \infty$$

Thus we must look for solutions of (7.71) for all x rather than just those x between 0 and a .

If we consider first the possibility of states with $E < 0$, the coefficient $-2mE/\hbar^2$ in front of ψ in (7.71) is positive and we can write (7.71) as

$$\psi''(x) = \alpha^2\psi(x)$$

where $\alpha = \sqrt{-2mE}/\hbar$. Just as with the rigid box, this equation has the solutions $e^{\alpha x}$ and $e^{-\alpha x}$ or any combination of both:

$$\psi(x) = Ae^{\alpha x} + Be^{-\alpha x} \quad (7.72)$$

But in the present case we can immediately see that none of these solutions can possibly be physically acceptable. The point is that (7.72) is the solution in the whole range $-\infty < x < \infty$. Now, as $x \rightarrow \infty$, the exponential $e^{\alpha x}$ grows without limit or “blows up,” and it is not physically reasonable to have a wave function $\psi(x)$ that grows without limit as we move farther from the origin. Such a $\psi(x)$ cannot be normalized. The only way out of this difficulty is to have the coefficient A of $e^{\alpha x}$ in (7.72) equal to zero. Similarly, as $x \rightarrow -\infty$, the exponential $e^{-\alpha x}$ blows up; thus by the same argument the coefficient B must also be zero, and we are left with just the zero solution $\psi(x) \equiv 0$. That is, there are no acceptable states with $E < 0$, just as we expected.

The argument just given crops up surprisingly often in solving the Schrödinger equation. If a solution of the equation blows up as $x \rightarrow \infty$ or as $x \rightarrow -\infty$, that solution is obviously not acceptable. Thus we can add to our list of conditions that must be satisfied by an acceptable wave function $\psi(x)$ the requirement that $\psi(x)$ must not blow up as $x \rightarrow \pm\infty$. We speak of a function that satisfies this requirement as being “well behaved” as $x \rightarrow \pm\infty$. This requirement is actually another example of a boundary condition, since the “points” $x = \pm\infty$ are the boundaries of our system.

Let us next examine the possibility of states of our free particle with $E \geq 0$. In this case the Schrödinger equation can be written as

$$\psi''(x) = -\left(\frac{2mE}{\hbar^2}\right)\psi(x) = -k^2\psi(x) \quad (7.73)$$

where

$$k = \frac{\sqrt{2mE}}{\hbar} \quad (7.74)$$

As before, the general solution of this equation is

$$\psi(x) = A \sin kx + B \cos kx \quad (7.75)$$

The important point about this solution is that neither $\sin kx$ nor $\cos kx$ blows up as $x \rightarrow \pm\infty$. Thus neither function suffers the difficulty that we encountered with negative energies,* and, for any value of k , the function (7.75) is an acceptable solution, for any constants A and B . According to (7.74), this means that all energies in the continuous range $0 \leq E < \infty$ are allowed. In particular, the energy of a free particle is not quantized. Evidently, it is only when a particle is confined in some way, that its energy is quantized.

To understand what the positive-energy wave functions (7.75) represent, it is helpful to recall the identities (Problem 7.16)

$$\sin kx = \frac{e^{ikx} - e^{-ikx}}{2i} \quad \text{and} \quad \cos kx = \frac{e^{ikx} + e^{-ikx}}{2} \quad (7.76)$$

Substituting these expansions into the wave function (7.75), we can write

$$\psi(x) = Ce^{ikx} + De^{-ikx} \quad (7.77)$$

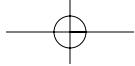
where you can easily find C and D in terms of the original coefficients A and B . It is important to note that since A and B were arbitrary, the same is true of C and D ; that is, (7.77) is an acceptable solution for any values of C and D .

The full, time-dependent wave function $\Psi(x, t)$ for the spatial function (7.77) is

$$\Psi(x, t) = \psi(x)e^{-i\omega t} = Ce^{i(kx-\omega t)} + De^{-i(kx+\omega t)} \quad (7.78)$$

This is a superposition of two traveling waves, one moving to the right (with coefficient C) and the other moving to the left (with coefficient D). If we choose the coefficient $D = 0$, then (7.78) represents a particle with definite momentum $\hbar k$ to the right; if we choose $C = 0$, then (7.78) represents a particle with momentum of the same magnitude $\hbar k$ but directed to the left. If both C and D are nonzero, then (7.78) represents a superposition of both momenta.

*Although the function (7.75) doesn't blow up as $x \rightarrow \pm\infty$, it does still suffer a lesser difficulty, that it cannot be normalized since $\int_{-\infty}^{\infty} |\psi(x)|^2 dx$ is infinite. However, this difficulty can be circumvented since we can build normalizable functions out of (7.75) using the Fourier integral.


224 Chapter 7 • The Schrödinger Equation in One Dimension

7.8 The Nonrigid Box

So far, our only example of a particle that is confined, or bound, is the rather unrealistic case of a particle in a perfectly rigid box, the infinite square well. In this section we apply the Schrödinger equation to a particle in the more realistic nonrigid box, a potential well of finite depth. This is a rather long section, but the ideas it contains are all fairly simple and are central to an understanding of many quantum systems.

The first step in applying the Schrödinger equation to any system is to determine the potential-energy function. Therefore, we must first decide what is the potential energy, $U(x)$, of a particle in a nonrigid box. For a *rigid* box we know that

$$U(x) = \begin{cases} 0 & 0 \leq x \leq a \\ \infty & x < 0 \text{ and } x > a \end{cases} \quad (7.79)$$

No finite amount of energy can remove the particle from a perfectly rigid box.* For most systems, a more realistic assumption would be that there is a *finite* minimum energy needed to remove a stationary particle from the box. If we call this minimum energy U_0 , the potential-energy function would be

$$U(x) = \begin{cases} 0 & 0 \leq x \leq a \\ U_0 & x < 0 \text{ and } x > a \end{cases} \quad (7.80)$$

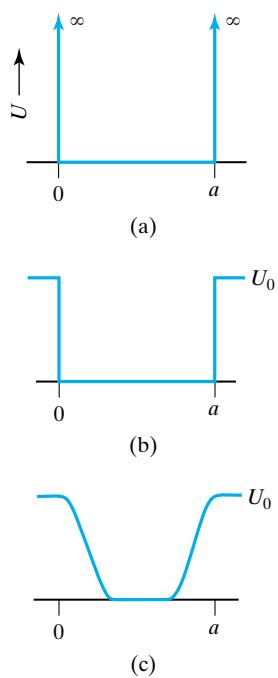


FIGURE 7.8

Three potential wells: (a) the infinite well (7.79); (b) the finite square well (7.80); (c) a finite rounded well.

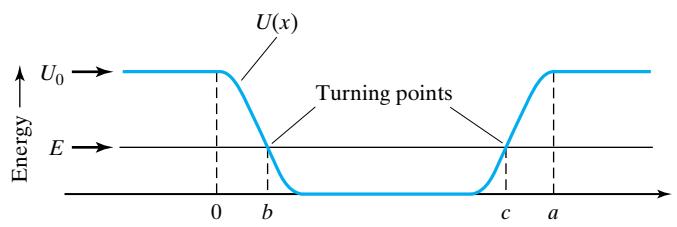
This potential, which we call the nonrigid box, is often called a finite square well. In Fig. 7.8(a) and (b) we plot the potential-energy functions (7.79) and (7.80).

Even the finite square well of Fig. 7.8(b) is somewhat unrealistic in that the potential energy jumps abruptly from 0 to U_0 at $x = 0$ and $x = a$. For a real particle in a box (for example, an electron in a conductor) the potential energy changes continuously near the walls, more like the well shown in Fig. 7.8(c). This well is sometimes called a *rounded well*. To simplify our discussion, we will suppose that the rounded well has $U(x)$ exactly constant, $U(x) = U_0$, for $x < 0$ and $x > a$, as shown in Fig. 7.8(c).

In this section we want to investigate the energy levels of a particle confined in a nonrigid box such as either Fig. 7.8(b) or (c). As one might expect, the properties of both wells are qualitatively similar.

Like the infinite well, the finite wells allow no states with $E < 0$ if we define the zero of U at the bottom of the well. (See Problem 7.38.) An important difference between the infinite and finite wells is that in the finite well the particle can escape from the well if $E > U_0$. This means that the wave functions for $E > U_0$ are quite similar to those of a free particle. In particular, the possible energies for $E > U_0$ are not quantized, but we will not pursue this point here since our main interest is in the bound states, whose energies lie in the interval $0 < E < U_0$.

*Until about 30 years ago, one would have said that in this respect the perfectly rigid box is totally unrealistic — a real bound system might require a *large* energy to pull it apart, but surely not an infinite amount. As we discuss in Chapter 18, we now know that subatomic particles like neutrons and protons are made up of sub-subatomic particles called quarks, and that an infinite energy is needed to pull them apart (that is, they cannot be pulled apart). Thus a potential energy function like (7.79) may be more realistic than we had formerly appreciated.

**FIGURE 7.9**

The classical turning points. A classical particle of energy E trapped in the potential well oscillates back and forth, turning around at the points $x = b$ and $x = c$, where the kinetic energy is zero and hence $E = U(x)$.

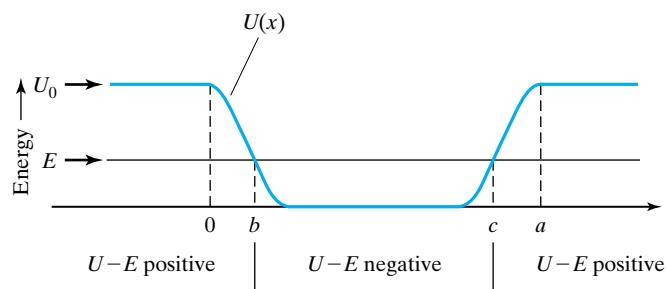
A classical particle moving in a finite well with energy in the interval $0 < E < U_0$ would simply bounce back and forth indefinitely. In the square well of Fig. 7.8(b) it would bounce between the points $x = 0$ and $x = a$. For the rounded well, the points at which the particle turns around are determined by the condition $E = U(x)$ (since the kinetic energy must be zero at the turning point where the particle comes instantaneously to rest). These points can be found graphically as in Fig. 7.9 by drawing a horizontal line at the height representing the energy E . The points $x = b$ and $x = c$, at which this line meets the potential-energy curve, are the two classical turning points, and a classical particle with energy E simply bounces back and forth between these points.

We now consider the Schrödinger equation,

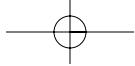
$$\psi''(x) = \frac{2m}{\hbar^2} [U(x) - E]\psi$$

for a quantum particle in a finite potential well. We seek values of E in the range $0 < E < U_0$, which possess physically acceptable solutions. To understand when we should expect to find allowed energies, it is useful to examine the general behavior of solutions of the Schrödinger equation.

Focusing attention on a particular value of E (with $0 < E < U_0$), we can distinguish two important ranges of x : those x where the factor $[U(x) - E]$ is positive and those x where it is negative. The dividing points between these regions are the classical turning points $x = b$ and $x = c$, where $U(x) = E$. (These were defined in Fig. 7.9 and are shown again in Fig. 7.10.) The regions where $[U(x) - E]$ is positive are outside these turning points ($x < b$ and $x > c$) and are often called the *classically forbidden regions* since a classical particle with energy E cannot penetrate there. The region where $[U(x) - E]$ is negative is the interval $b < x < c$ and is called the classically allowed region. The behavior of the wave function $\psi(x)$ is quite different in these two regions.

**FIGURE 7.10**

The factor $[U(x) - E]$, which appears on the right side of the Schrödinger equation, is positive for $x < b$ and $x > c$ and is negative for $b < x < c$.



226 Chapter 7 • The Schrödinger Equation in One Dimension

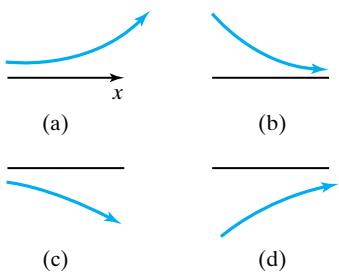


FIGURE 7.11

If $\psi(x)$ satisfies an equation of the form (7.81), it is concave away from the axis.

Wave Functions Outside the Well

In the region where $[U(x) - E]$ is positive, the Schrödinger equation has the form

$$\psi''(x) = (\text{positive function}) \times \psi(x) \quad (7.81)$$

where the “positive function” is $(2m/\hbar^2)[U(x) - E]$. In an interval where $\psi(x)$ is positive, this implies that $\psi''(x)$ is also positive and hence that $\psi'(x)$ is increasing and $\psi(x)$ is concave *upward*, as in Fig. 7.11(a) and (b). If $\psi(x)$ is negative, then (7.81) implies that $\psi''(x)$ is negative and hence that $\psi(x)$ is concave *downward*, as in Fig. 7.11(c) and (d). In either case we see that $\psi(x)$ is concave *away from the x-axis*.

The behavior shown in Fig. 7.11 can be seen explicitly if we look in either of the regions $x < 0$ or $x > a$, where $U(x) = U_0 = \text{constant}$. The explicit form of $\psi(x)$ is readily found by solving the Schrödinger equation. Since $(U_0 - E) > 0$, we can define a number α by the equation

$$\frac{2m}{\hbar^2}[U_0 - E] = +\alpha^2 \quad (7.82)$$

and the Schrödinger equation becomes

$$\psi''(x) = +\alpha^2\psi(x) \quad (7.83)$$

As we have seen before, the general solution of this differential equation has the form

$$\psi(x) = Ae^{\alpha x} + Be^{-\alpha x} \quad (7.84)$$

with A and B arbitrary. As you can easily check, any function of this form is concave away from the x -axis as in Fig. 7.11.

If we look on the left of the well ($x < 0$), then, as $x \rightarrow -\infty$, the exponential $e^{-\alpha x}$ blows up and is physically unacceptable. Thus, in the region $x < 0$, the solution (7.84) is physically acceptable only if $B = 0$.

Similarly, in the region $x > a$, the general solution has the same form

$$\psi(x) = Ce^{\alpha x} + De^{-\alpha x} \quad x > a \quad (7.85)$$

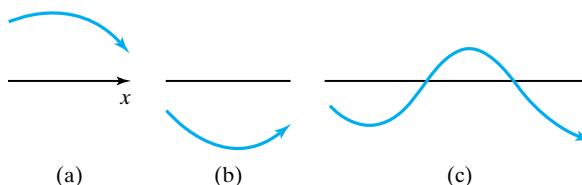
and this is acceptable only if $C = 0$. Thus, in these classically forbidden regions the physically acceptable wave functions die away exponentially as x goes to $\pm\infty$.

Wave Functions within the Well

In the region $b < x < c$, where $U(x) < E$, the Schrödinger equation has the form

$$\psi''(x) = (\text{negative function}) \times \psi(x) \quad (7.86)$$

we can argue that $\psi(x)$ is concave *toward* the axis and tends to oscillate, as follows: If $\psi(x)$ is positive, then $\psi''(x)$ is negative and $\psi(x)$ curves *downward*, as in Fig. 7.12(a); if $\psi(x)$ is negative, the argument reverses and $\psi(x)$ bends

**FIGURE 7.12**

If $\psi(x)$ satisfies an equation of the form (7.86), it curves toward the axis and tends to oscillate.

upward, as in Fig. 7.12(b). In either case, $\psi(x)$ curves toward the axis. If the interval $b < x < c$ is sufficiently wide, a function bending toward the axis will cross the axis and immediately start bending the other way, as in Fig. 7.12(c). Thus we can say that in the region $b < x < c$ the wave function tends to oscillate about the axis.

If the negative function in (7.86) has a large magnitude, then $\psi''(x)$ tends to be large and $\psi(x)$ curves and oscillates rapidly. Conversely, when the negative function is small, $\psi(x)$ bends gradually and oscillates slowly. This is all physically reasonable: The negative function in (7.86) is proportional to $[U(x) - E]$, which is just the negative of the kinetic energy; according to de Broglie, large kinetic energy means short wavelength and hence rapid oscillation, and vice versa.

For the case of a finite square well, where $U_0 = \text{constant} = 0$ within the well, the negative function of (7.86) is $-2mE/\hbar^2$. We can define a constant k by the equation

$$-\frac{2m}{\hbar^2}E = -k^2 \quad (7.87)$$

and the Schrödinger equation becomes

$$\psi''(x) = -k^2\psi \quad (7.88)$$

The most general solution of this equation is

$$\psi(x) = F \sin kx + G \cos kx \quad (7.89)$$

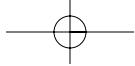
where F and G are arbitrary constants.

Solving for the allowed energies of a particle in a finite square well is rather messy. It involves starting with the general forms (7.84), (7.85), and (7.89), and then using the conditions that $\psi(x)$ and $\psi'(x)$ must be continuous at the edges of the well to solve for the coefficients A, D, F , and G (B and C are already known to be zero). It turns out there is no simple analytic solution to this problem and a numerical solution is needed, as explored in Problem 7.67. However, as we show in the next subsection, one can determine the detailed qualitative behavior of the solutions without doing any calculations at all.

Searching for Allowed Energies

Now that we understand the qualitative behavior of solutions of the Schrödinger equation, let us return to our hunt for acceptable solutions. We consider the most general case of a rounded well of finite depth, and we start in the region $x < 0$, where $U(x)$ is constant (Fig. 7.10). There, the known, acceptable form is

$$\psi(x) = Ae^{\alpha x} \quad (x < 0)$$



228 Chapter 7 • The Schrödinger Equation in One Dimension

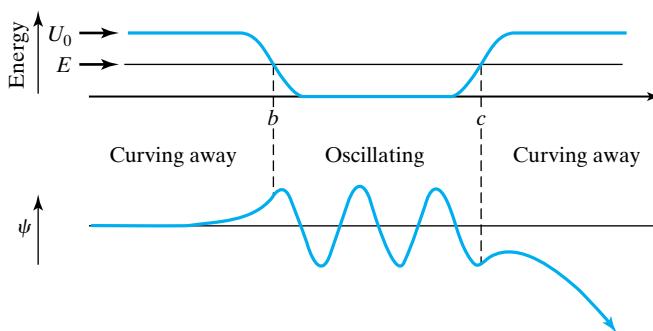


FIGURE 7.13

The wave function oscillates between the two turning points $x = b$ and $x = c$ and curves away from the axis outside them. The example shown is well behaved as $x \rightarrow -\infty$, but blows up as $x \rightarrow +\infty$.

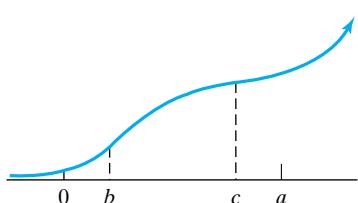


FIGURE 7.14

When E is very small, the wave function bends too slowly inside the well. The function that has the form $Ae^{\alpha x}$ when $x < 0$ blows up as $x \rightarrow +\infty$.

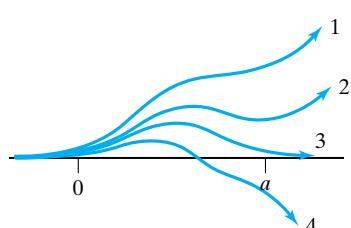


FIGURE 7.15

Solutions of the Schrödinger equation for four successively larger energies. All four solutions have the well-behaved form $Ae^{\alpha x}$ for $x < 0$, but only number 3 is also well behaved as $x \rightarrow +\infty$.

FIGURE 7.16

Wave functions for the second and third energy levels of a nonrigid box.

When we move to the right, our solution will cease to have this explicit form once $U(x)$ starts to vary, but it will continue to bend away from the axis until x reaches the point b . At $x = b$, it will start oscillating and continue to do so until $x = c$, where it will start curving away from the axis again. Thus, the general appearance of $\psi(x)$ will be as shown in Fig. 7.13, with two regions where $\psi(x)$ bends away from the axis, separated by one region where $\psi(x)$ oscillates. Figure 7.13 shows a solution that blows up on the right. We must now find out if there are any values of E for which the solution is well behaved both on the left and right.

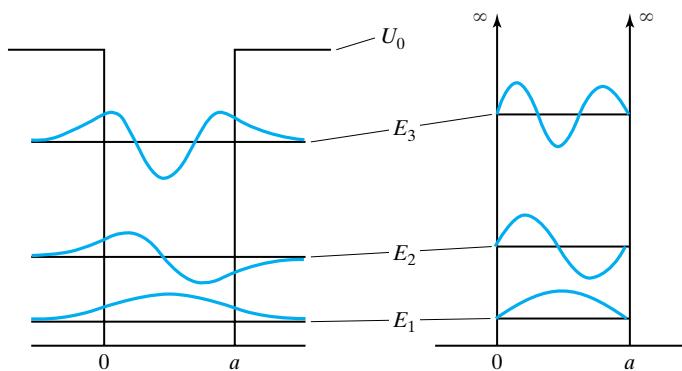
Let us begin a systematic search for allowed energies, starting with E close to zero. We will show first that with E sufficiently close to zero, an acceptable wave function is impossible. We start with the acceptable wave function $Ae^{\alpha x}$ in the region $x < 0$, and follow $\psi(x)$ to the right. When we reach the classical turning point $x = b$, $\psi(x)$ has a positive slope and starts to bend toward the axis. But with E very small, $\psi(x)$ bends very slowly. Thus, when we reach the second turning point $x = c$, the slope is still positive. With ψ and ψ' both positive, $\psi(x)$ continues to increase without limit, as shown in Fig. 7.14. Therefore, the wave function which is well behaved as $x \rightarrow -\infty$ blows up as $x \rightarrow +\infty$, and we conclude that there can be no acceptable wave function with E too close to zero.

Suppose now that we slowly increase E , continuing to hunt for an allowed energy. For larger values of E , the kinetic energy is larger and, as we have seen, $\psi(x)$ bends more rapidly inside the well. Thus it can bend enough that its slope becomes negative, as shown in Fig. 7.15, curve 2. Eventually, its value and slope at the right of the well will be just right to join onto the solution that is well behaved as $x \rightarrow \infty$, and we have an acceptable wave function (curve 3 in Fig. 7.15). If we increase E any further, $\psi(x)$ will bend over too far inside the well and will now approach $-\infty$ as $x \rightarrow \infty$, like curve 4 in Fig. 7.15. Evidently, there is exactly one allowed energy in the range explored so far.

If E is increased still further, the wave function may bend over and back just enough to fit onto the function that is well behaved as $x \rightarrow \infty$ as in Fig. 7.16(a). If this happens, we have a second allowed energy. Beyond this, we may find a third acceptable wave function, like that in Fig. 7.16(b), and so on.

Figure 7.17 shows the first three wave functions for a finite square well, beside the corresponding wave functions for an infinitely deep square well of



**FIGURE 7.17**

The lowest three energy levels and wave functions for a finite square well and for an infinite square well of the same width. The horizontal lines that represent each energy level have been used as the axes for drawing the corresponding wave functions.

the same width a . Notice the marked similarity of corresponding functions: The first wave function of the infinite well fits exactly half an oscillation into the well, while that of the finite well fits somewhat less than half an oscillation into the well since it doesn't actually vanish until $x = \pm\infty$. Similarly, the second function of the infinite well makes one complete oscillation, whereas that of the finite well makes just less than one oscillation. For this reason, the energy of each level in the finite well is slightly lower than that of the corresponding level in the infinite well.

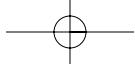
It is useful to note that the ground-state wave function for any finite well has no nodes, while that for the second level has one node, and that for the n th level has $n - 1$ nodes. This general trend (more nodes for higher energies) could have been anticipated since higher energy corresponds to a wave function that oscillates more quickly and hence has more nodes.

The wave functions in Fig. 7.17 illustrate two more important points. First, the wave functions of the finite well are nonzero outside the well, in the classically forbidden region. (Remember that a classical particle with energy $0 < E < U_0$ cannot escape outside the turning points.) However, since the wave function is largest inside the well, the particle is most likely to be found inside the well; and since $\psi(x)$ approaches zero rapidly as one moves away from the well, we can say that our particle is bound inside, or close to, the potential well. Nevertheless, there is a definite nonzero probability of finding the particle in the classically forbidden regions. This difference between classical and quantum mechanics is due to the wave nature of quantum particles. The ability of the quantum wave function to penetrate classically forbidden regions has important consequences, as we discuss in Section 7.10

A second important point concerns the number of bound states of the finite well. With the infinite well, one can increase E indefinitely and always encounter more bound states. With the finite well, however, the particle is no longer confined when E reaches U_0 , and there are no more bound states. The number of bound states depends on the well depth U_0 and width a , but it is always finite. The number of bound states can be approximately computed using a simple method described in Problem 7.45.

7.9 The Simple Harmonic Oscillator*

*The simple harmonic oscillator plays an amazingly important role in many areas of quantum physics. Nevertheless, we won't be using the results of this section again until Chapter 12 on molecules. Thus you could, if you wish, omit it on a first reading.



230 Chapter 7 • The Schrödinger Equation in One Dimension

As another example of a one-dimensional bound particle, we consider the simple harmonic oscillator. The name **simple harmonic oscillator (or SHO)** is used in classical and quantum mechanics for a system that oscillates about a stable equilibrium point to which it is bound by a force obeying Hooke's law, $F = -kx$.

Familiar classical examples of harmonic oscillators are a mass suspended from an ideal spring and a pendulum oscillating with small amplitude. Before we give any quantum examples, it may be worth recalling why the harmonic oscillator is such an important system. If a particle is in equilibrium at a point x_0 , the total force F on the particle is zero at x_0 ; that is, $F(x_0) = 0$. If the particle is displaced slightly to a neighboring point x , the total force will be approximately

$$F(x) = F(x_0) + F'(x_0)(x - x_0) \quad (7.90)$$

with $F(x_0) = 0$ in this case. If x_0 is a point of stable equilibrium, the force is a restoring force; that is, $F(x)$ is negative when $x - x_0$ is positive, and vice versa. Therefore, $F'(x_0)$ must be negative and we denote it by $-k$, where k is called the *force constant*. With $F'(x_0) = -k$ and $F(x_0) = 0$, Eq. (7.90) becomes

$$F(x) = -k(x - x_0) \quad (7.91)$$

which is Hooke's law. Thus, any particle oscillating about a stable equilibrium point will oscillate harmonically* for sufficiently small displacements ($x - x_0$).

An important example of a quantum harmonic oscillator is the motion of any one atom inside a solid crystal; each atom has a stable equilibrium position relative to its neighboring atoms and can oscillate harmonically about that position. Another important example is a diatomic molecule, such as HCl, whose two atoms can vibrate harmonically, in and out from one another.

In quantum mechanics we work, not with the force F , but with the potential energy U . This is easily found by integrating (7.91) to give

$$U(x) = - \int_{x_0}^x F(x) dx = \frac{1}{2}k(x - x_0)^2 \quad (7.92)$$

if we take U to be zero at x_0 . Thus, in quantum mechanics the SHO can be characterized as a system whose potential energy has the form (7.92). This function is a parabola, with its minimum at $x = x_0$, as shown in Fig. 7.18(a).

It is important to remember that (7.91) and (7.92) are approximations that are usually valid only for small displacements from x_0 . This point is illustrated in Fig. 7.18(b), which shows the potential energy of a typical diatomic molecule, as a function of the distance r between the two atoms. The molecule is in equilibrium at the separation r_0 . For r close to r_0 , the potential energy is well approximated by a parabola of the form $U(r) = \frac{1}{2}k(r - r_0)^2$, but when r is far from r_0 , $U(r)$ is quite different. Thus for small displacements the molecule will behave like an SHO, but for large displacements it will not. This same statement can be made about almost any oscillating system. (For example,

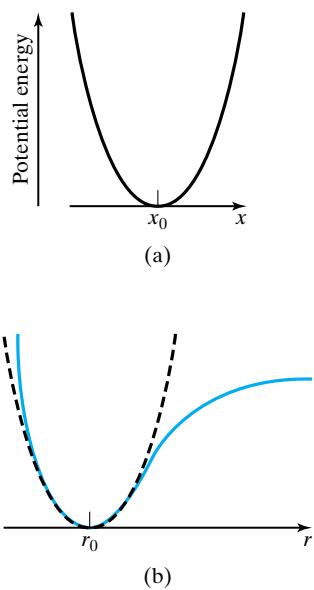


FIGURE 7.18

(a) The potential energy of an ideal simple harmonic oscillator is a parabola. (b) The potential energy of a typical diatomic molecule (solid curve) is well approximated by that of an SHO (dashed curve) when r is close to its equilibrium value r_0 .

*Unless, of course, the first derivative happens to vanish at the equilibrium point, $F'(x_0) = 0$. This occurs for the quartic potential $U(x) \propto x^4$, with $F(x) \propto -x^3$, but we will not consider this special case here.

Section 7.9 • The Simple Harmonic Oscillator **231**

simple pendulum is well known to oscillate harmonically for small amplitudes, but not when the amplitude is large.) It is because small displacements from equilibrium are very common that the harmonic oscillator is so important.

There is a close connection between the classical and quantum harmonic oscillators, and we start with a quick review of the former. If we choose our origin at the equilibrium position, then $x_0 = 0$ and the force is $F = -kx$. If the particle has mass m , Newton's second law reads

$$ma = -kx$$

If we define the classical angular frequency

$$\omega_c = \sqrt{\frac{k}{m}} \quad (7.93)$$

this becomes

$$\frac{d^2x}{dt^2} = -\omega_c^2 x$$

which has the general solution

$$x = a \sin \omega_c t + b \cos \omega_c t \quad (7.94)$$

Thus, the position of the classical SHO varies sinusoidally in time with angular frequency ω_c . If we choose our origin of time, $t = 0$, at the moment when the particle is at $x = 0$ and moving to the right, then (7.94) takes the form

$$x = a \sin \omega_c t \quad (7.95)$$

The positive number a is the amplitude of the oscillations, and the particle oscillates between $x = a$ and $x = -a$. In other words, the points $x = \pm a$ are the classical turning points. When the particle is at $x = a$, all of its energy is potential energy; thus, $E = \frac{1}{2}ka^2$ and hence

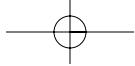
$$a = \sqrt{\frac{2E}{k}} \quad (7.96)$$

As one would expect, the classical amplitude a increases with increasing energy.

To find the allowed energies of a *quantum* harmonic oscillator, we must solve the Schrödinger equation with $U(x)$ given by

$$U(x) = \frac{1}{2}kx^2 \quad (7.97)$$

where we have again chosen $x_0 = 0$. Qualitatively, the analysis is very similar to that for the finite well discussed in the preceding section. For any choice of E , the solutions will bend away from the axis outside the classical turning points $x = a$ and $x = -a$ and will oscillate between these points. As before, most values of E do not produce an acceptable solution; only special values of E produce a solution that satisfies the boundary conditions, that is, decays exponentially as $x \rightarrow -\infty$ and $x \rightarrow \infty$. Thus, the allowed energies are quantized. The potential energy (7.97) increases indefinitely as x moves away from the



232 Chapter 7 • The Schrödinger Equation in One Dimension

origin [see Fig. 7.18(a)], and the particle is therefore confined for all energies. In this respect, the SHO resembles the infinitely deep potential well, and we would expect to find infinitely many allowed energies, all of them quantized.

An important difference between the SHO and most other potential wells is that the Schrödinger equation for the SHO can be solved analytically. The solution is quite complicated and will not be given here. However, the answer for the energy levels is remarkably simple: The allowed energies turn out to be

$$E = \frac{1}{2}\hbar\sqrt{\frac{k}{m}}, \quad \frac{3}{2}\hbar\sqrt{\frac{k}{m}}, \quad \frac{5}{2}\hbar\sqrt{\frac{k}{m}}, \dots$$

The quantity $\sqrt{k/m}$ is the frequency ω_c , defined in (7.93), of a classical oscillator with the same force constant and mass. It is traditional to rewrite the allowed energies in terms of ω_c , as

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_c \quad n = 0, 1, 2, \dots \quad (7.98)$$

As anticipated, the allowed energies are all quantized, and there are levels with arbitrarily large energies. A remarkable feature of (7.98) is that the energy levels of the harmonic oscillator are all equally spaced. This is a property that we could not have anticipated in our qualitative discussion, but which emerges from the exact solution.

The allowed energies of the harmonic oscillator are shown in Fig. 7.19, which also shows the corresponding wave functions (each drawn on the line representing its energy). Notice the similarity of these wave functions to those of the finite well shown in Fig. 7.17. In particular, just as with the finite well, the lowest wave function has no nodes, the next has one node, and so on. Notice also that the wave functions with higher energy spread out farther from $x = 0$, just as the classical turning points $x = \pm a$ move farther out when E increases. Finally, note that the ground state ($n = 0$) has a zero-point energy (equal to $\frac{1}{2}\hbar\omega_c$) as required by the uncertainty principle, as discussed in connection with Eq. (6.39).

The wave functions shown in Fig. 7.19 were plotted using the known analytic solutions, which we list in Table 7.1, where, for convenience, we have introduced the parameter

$$b = \sqrt{\frac{\hbar}{m\omega_c}} \quad (7.99)$$

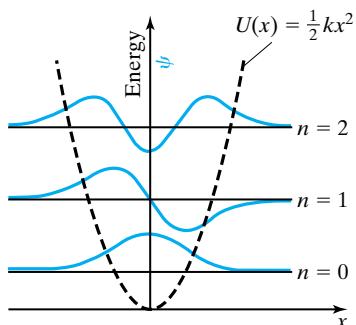


FIGURE 7.19

The first three energy levels and wave functions of the simple harmonic oscillator.

TABLE 7.1

The energies and wave functions of the first three levels of a quantum harmonic oscillator. The length b is defined as $\sqrt{\hbar/m\omega_c}$.

n	E_n	$\psi(x)$
0	$\frac{1}{2}\hbar\omega_c$	$A_0 e^{-x^2/2b^2}$
1	$\frac{3}{2}\hbar\omega_c$	$A_1 \frac{x}{b} e^{-x^2/2b^2}$
2	$\frac{5}{2}\hbar\omega_c$	$A_2 \left(1 - 2\frac{x^2}{b^2}\right) e^{-x^2/2b^2}$

(This parameter has the dimensions of a length and is, in fact, the half-width of the SHO well at the ground-state energy — see Problem 7.48.) The factors A_0 , A_1 , and A_2 in the table are normalization constants.*

Example 7.4

Verify that the $n = 1$ wave function given in Table 7.1 is a solution of the Schrödinger equation with $E = \frac{3}{2}\hbar\omega_c$. (For the cases $n = 0$ and $n = 2$, see Problems 7.49 and 7.52.)

The potential energy is $U = \frac{1}{2}kx^2$, and the Schrödinger equation is therefore

$$\psi'' = \frac{2m}{\hbar^2} \left(\frac{1}{2}kx^2 - E \right) \psi \quad (7.100)$$

Differentiating the wave function ψ_1 , of Table 7.1, we find that

$$\psi'_1 = A_1 \left(\frac{1}{b} - \frac{x^2}{b^3} \right) e^{-x^2/2b^2}$$

and

$$\psi''_1 = A_1 \left(-\frac{3x}{b^3} + \frac{x^3}{b^5} \right) e^{-x^2/2b^2} = \left(\frac{x^2}{b^4} - \frac{3}{b^2} \right) A_1 \frac{x}{b} e^{-x^2/2b^2}$$

where the last three factors together are just ψ_1 . Using (7.99) to replace b , we find

$$\psi''_1 = \left(\frac{m^2\omega_c^2 x^2}{\hbar^2} - \frac{3m\omega_c}{\hbar} \right) \psi_1$$

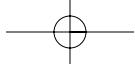
Replacing ω_c^2 by k/m in the first term, we get

$$\psi''_1 = \frac{2m}{\hbar^2} \left(\frac{1}{2}kx^2 - \frac{3}{2}\hbar\omega_c \right) \psi_1$$

which is precisely the Schrödinger equation (7.100), with $E = \frac{3}{2}\hbar\omega_c$ as claimed.

The properties of the quantum SHO are nicely illustrated by the example of a diatomic molecule, such as HCl. A diatomic molecule is, of course, a complicated, three-dimensional system. However, the energy of the molecule can be expressed as the sum of three terms: an electronic term, corresponding to the motions of the individual electrons; a rotational term, corresponding to rotation of the whole molecule; and a vibrational term, corresponding to the in-and-out, radial vibrations of the two atoms. Careful analysis of molecular spectra lets one disentangle the possible values of these three terms. In

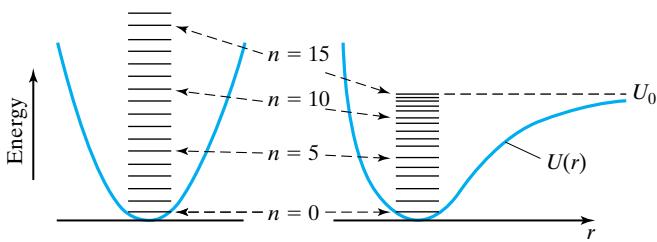
*The three functions in Table 7.1 illustrate what is true for all n , that the wave function of the n th level has the form $P_n(x) \exp(-x^2/2b^2)$, where $P_n(x)$ is a certain polynomial of degree n , called a Hermite polynomial.



234 Chapter 7 • The Schrödinger Equation in One Dimension

FIGURE 7.20

Vibrational levels of a typical diatomic molecule (right) and the SHO that approximates the molecule for small displacements (left). The first five or so levels correspond very closely; the higher levels of the molecule are somewhat closer together, and the molecule has no levels above the energy U_0 .



In particular, the vibrational motion of the two atoms is one-dimensional (along the line joining the atoms) and the potential energy approximates the SHO potential. Therefore, the allowed values of the vibrational energy should approximate those of an SHO, and this is amply borne out by observation, as illustrated in Fig. 7.20.

The molecular potential energy $U(r)$ in Fig. 7.20 deviates from the parabolic shape of the SHO at higher energies. Therefore, we would expect the higher energy levels to depart from the uniform spacing of the SHO. This, too, is what is observed.

The observation of photons emitted when molecules make transitions between different vibrational levels is an important source of information on interatomic forces, as we discuss further in Chapter 12. These same photons can also be used to identify the molecule that emitted them. For example, the H_2 molecule emits infrared photons of frequency $1.2 \times 10^{14} \text{ Hz}$ when it drops from one vibrational level to the next. This radiation is used by astronomers to locate clouds of H_2 molecules in our galaxy.

7.10 Tunneling*

*The phenomenon of tunneling is perhaps second only to quantization of energy as a uniquely quantum phenomenon. It has many important applications, as we will mention here and in Chapters 14 and 17. However, we will not be using this material again until Chapter 14, so you could omit this section for now.

So far in this chapter, we have focused mainly on a particle that is confined inside a potential well, such as the rigid box of Section 7.6, or the harmonic oscillator well of Section 7.9. We have seen that for these confined, or bound, particles, the Schrödinger equation implies quantization of the allowed energies — one of the most dramatic differences between classical and quantum mechanics. We have also discussed one example of an unconfined particle, the free particle of Section 7.7, for which we found that the Schrödinger equation does not imply any quantization of energy. In this section we discuss a second example of an unconfined particle, for which the Schrödinger equation has implications almost as dramatic as the quantization of energy for bound states. Specifically, we will consider a particle whose potential energy has a “barrier” (to be defined in a moment). In classical mechanics such barriers are impenetrable and a particle on either side of the barrier cannot cross over to the other side. In quantum mechanics we will find that the particle can “tunnel” through the barrier and emerge on the other side. This **barrier penetration**, or **tunneling**, has dramatic consequences in several natural phenomena, such as radioactive decay (Chapter 17), and is the basis of several modern electronic devices, such as the scanning tunneling microscope (Chapter 14).

A simple example of a barrier is shown in Fig. 7.21, which is a plot of the potential energy of an electron moving along an x -axis consisting of two identical conducting wires separated by a small gap from x_0 to x_1 . The gap between

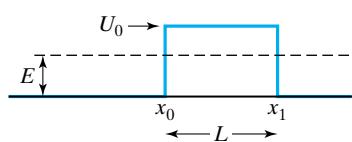
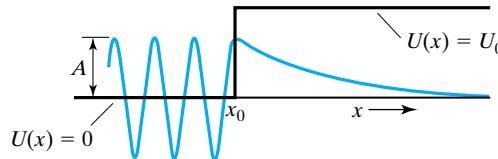


FIGURE 7.21

A rectangular barrier of height U_0 and width L .



the two conductors could be just an air gap, or it could be a thin layer of dirt in a poorly made electrical connection. In either case the gap forms a barrier of thickness $L = x_1 - x_0$ between the two wires. Inside either wire the potential energy is a constant, which we can take to be zero, but in the barrier it has a higher value $U_0 > 0$. We will be interested in the case that the particle's energy is less than the barrier height, $0 < E < U_0$, as shown in Fig. 7.21. Under these conditions a classical particle is excluded from the barrier since with $E < U_0$ its kinetic energy between x_0 and x_1 would be negative — an impossibility in classical mechanics. Thus, a classical particle could approach the barrier from the left (for example) but would inevitably rebound straight back on arrival at x_0 ; it could certainly not emerge on the right of x_1 unless we gave it more energy. To see in detail what happens to a quantum particle when it hits this barrier, one must solve the Schrödinger equation with this potential-energy function $U(x)$. Although the solution is not especially difficult, we do not need to go through it since we can already understand its main features from our discussion of the finite square well in Section 7.8.

Consider first a barrier whose length L is *infinite*. This extreme case is shown in Fig. 7.22. It is sometimes called a *potential step* and is precisely the same as the right wall of the finite square well of Section 7.8 (Fig. 7.17). In Fig. 7.22 we have also shown the wave function $\psi(x)$ for an energy $E < U_0$. To the left of x_0 , the kinetic energy $K = (E - U)$ is positive and $\psi(x)$ is an oscillating sinusoidal wave, some combination of $\sin kx$ and $\cos kx$. To the right of x_0 , $(E - U)$ is negative and $\psi(x)$ is a decreasing exponential with the form

$$\psi(x) = Be^{-\alpha x} \quad (7.101)$$

where, you may recall [see (7.82) if you don't],

$$\alpha = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}} \quad (7.102)$$

As we emphasized in Section 7.8, the quantum particle has a nonzero probability of being found in the classically forbidden region where $E - U$ is negative.

In the infinitely long barrier of Fig. 7.22, $\psi(x)$ goes steadily to zero as x increases. But if the barrier has finite length, the situation is as shown in Fig. 7.23. Just as in Fig. 7.22, $\psi(x)$ is sinusoidal (with amplitude A_L) to the left

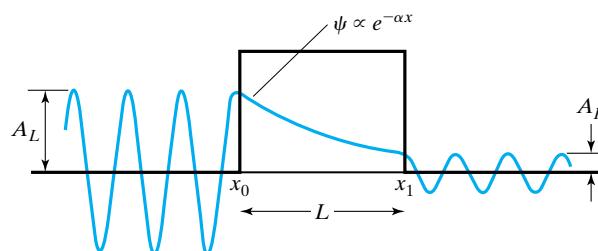
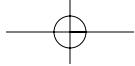


FIGURE 7.23

Wave function for a barrier of finite length. On the left ($x < x_0$), $\psi(x)$ is sinusoidal, with amplitude A_L ; in the barrier it decreases exponentially; on the right ($x > x_1$), it is sinusoidal again, with amplitude A_R .



236 Chapter 7 • The Schrödinger Equation in One Dimension

of x_0 ; and just as in Fig. 7.22, it decreases exponentially within the barrier ($x_0 < x < x_1$). But when we reach x_1 the barrier stops and, once again, $(E - U)$ is positive. Therefore, before $\psi(x)$ has decreased to zero, it starts oscillating again with amplitude A_R . The probability that the particle is to the left of x_0 , approaching the barrier, is proportional to A_L^2 ; the probability that it is to the right of the barrier is proportional to A_R^2 . Therefore, there is a nonzero probability P that a particle striking the barrier from the left will escape to the right*:

$$P = \left(\frac{A_R}{A_L} \right)^2 \quad (7.103)$$

From Fig. 7.23, we see that A_R is less than A_L because of the exponential decrease of $\psi(x)$ within the barrier. Specifically†

$$\frac{A_R}{A_L} \approx \frac{e^{-\alpha x_1}}{e^{-\alpha x_0}} = e^{-\alpha L} \quad (7.104)$$

Therefore, the probability that a particle which strikes the rectangular barrier of Fig. 7.23 with energy $E < U_0$ will tunnel through and emerge on the other side is

$$P \approx e^{-2\alpha L} \quad (7.105)$$

where α is given by (7.102).

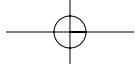
The probability (7.105) that a quantum particle will tunnel through the classically impenetrable barrier depends on the two variables: α , as given by (7.102), and L , the thickness of the barrier. In many applications the product αL is very large and the probability (7.105) is therefore very small. Nevertheless, if the particle keeps bumping against the barrier, it will eventually pass through it. This is what happens in the α decay of certain radioactive nuclei, as we will see in Section 17.10.

A modern application of quantum tunneling is the scanning tunneling microscope (STM). Here the surface of a sample is explored by measuring the electric current between the surface and a conducting probe that is scanned across the surface, just above it at a fixed height. This current is possible only because of quantum tunneling and the magnitude of the current is proportional to the probability (7.105) of tunneling through the barrier. This probability is sensitively dependent on L , the barrier thickness, that is, the distance between the probe and the surface. This means that by measuring the current, one can get a sensitive picture of the surface's topography, as we describe in detail in Chapter 14. (Meanwhile, see the beautiful picture of the surface of graphite made with an STM in Fig. 13.6.)

There are many other examples of quantum tunneling, but the two mentioned here are among the most important: The first is historically important

*Our discussion here is a little oversimplified. Strictly speaking, we should be using wave functions e^{ikx} that travel in a definite direction, rather than sinusoidal functions. Nevertheless, the result (7.103) is a good enough approximation.

†Here, too, we are oversimplifying. In Fig. 7.23 you can see that the value of the exponential wave function at x_0 is not quite the same as the amplitude A_L . (Similarly, the value at x_1 is not quite equal to A_R .) Nevertheless, under the conditions of interest, (7.104) is a satisfactory approximation.

Section 7.11 • The Time-Dependent Schrödinger Equation **237**

as the first application of quantum mechanics to a subatomic process, and the second is the basis for a modern and widely used experimental technique. Curiously, barrier penetration will not play an important role in our further development of quantum theory, and our focus for the next several chapters will be on bound states similar to those described in Sections 7.4 through 7.9. However, we will return to barrier penetration in Chapters 14 and 17.

Example 7.5

Consider two identical conducting wires, lying on the x axis and separated by an air gap of thickness $L = 1 \text{ nm}$, (that is, a few atomic diameters). An electron that is moving inside either conductor has potential energy zero, whereas in the gap its potential energy is $U_0 > 0$. Thus the gap is a barrier of the type illustrated in Fig. 7.21. The electron approaches the barrier from the left with energy E such that $U_0 - E = 1 \text{ eV}$; that is, the electron is 1 eV below the top of the barrier. What is the probability that the electron will emerge on the other side of the barrier? How different would this be if the barrier were twice as wide?

The required probability is given by Equation (7.105) with

$$\alpha = \frac{\sqrt{2m(U_0 - E)}}{\hbar} = \frac{\sqrt{2mc^2(U_0 - E)}}{\hbar c}$$

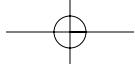
$$\approx \frac{\sqrt{2 \times (5 \times 10^5 \text{ eV}) \times (1 \text{ eV})}}{197 \text{ eV} \cdot \text{nm}} \approx 5.1 \text{ nm}^{-1}$$

Thus $\alpha L = 5.1$ and the transmission probability is $P = e^{-2\alpha L} = e^{-10.2} = 3.7 \times 10^{-5}$ or about 0.004%. This probability is not large, but if we send enough electrons at the barrier, some will certainly get through. If we double L , this will give a transmission probability $P' = e^{-4\alpha L} = e^{-20.4} = 1.4 \times 10^{-9}$ — a dramatically smaller result. This illustrates the extreme sensitivity of the transmission probability to the width of the gap — a property that is exploited in the scanning tunneling microscope.

7.11 The Time-Dependent Schrödinger Equation*

*The time-dependent Schrödinger equation is one of the principal cornerstones of quantum theory. Nevertheless, it is used surprisingly little in practical calculations, and we will use it only in Chapter 11. Thus, you could skip this section on a first reading without significant loss of continuity.

As we mentioned at the start of this chapter, there are in fact two Schrödinger equations — the time-independent Schrödinger equation, which has been our main subject up to now, and the time-*dependent* Schrödinger equation. The latter is one of the basic equations of quantum mechanics: It determines the time evolution of any quantum system and is the analog of Newton's second law, $\mathbf{F} = m\mathbf{a}$, in classical mechanics. Among all the many solutions of the time-dependent Schrödinger equation, the most important are the stationary states (or states of definite energy); and for these, the time-dependent Schrödinger equation reduces to the time-independent equation, as we will prove in this section. Because the stationary states are by far the most important states in practice, it is the time-independent Schrödinger equation that one uses in



238 Chapter 7 • The Schrödinger Equation in One Dimension

almost all practical problems. Nevertheless, you should have some acquaintance with the time-dependent equation, and that is the purpose of this section.

Before we write down the time-dependent Schrödinger equation, it is helpful to note that the time-independent Schrödinger equation (7.39) can be rewritten (as you should check) in the form

$$E\psi(x) = \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x) \right) \psi(x) \quad (7.106)$$

This is the equation that (we have claimed) determines the possible energies E and the corresponding spatial wave functions $\psi(x)$ for a particle whose potential energy is $U(x)$. The full wave function for the same particle, including its time dependence, is denoted $\Psi(x, t)$ and the time dependence of any such wave function is determined by the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t} \Psi(x, t) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right) \Psi(x, t) \quad (7.107)$$

There is no way to *prove* that this equation is correct. Just like its counterpart $\mathbf{F} = m\mathbf{a}$ in classical mechanics, it is ultimately an axiom, an assumption, that is justified by the success of its many predictions. Nevertheless, we can motivate the form of this equation with the following plausibility argument.

We are searching for a wave equation, a differential equation that wave functions must obey, and we can use the free-particle wave function as a test case. We assume that a free particle has the wave function $\Psi(x, t) = e^{i(kx - \omega t)}$ and has a (purely kinetic) energy given by $E = \hbar\omega = \hbar^2 k^2 / 2m$. Any candidate wave equation must have this wave function as its solution for the special case $U(x) = 0$. Guided by the energy equation $\hbar\omega = \hbar^2 k^2 / 2m$, we seek differential operators that produce factors of ω and k^2 , and we find that the operators $\partial/\partial t$ and $\partial^2/\partial x^2$ do just that:

$$\frac{\partial}{\partial t} e^{i(kx - \omega t)} = -i\omega e^{i(kx - \omega t)} \quad \text{and} \quad \frac{\partial^2}{\partial x^2} e^{i(kx - \omega t)} = -k^2 e^{i(kx - \omega t)}$$

And so we can construct an equation,

$$-i\hbar \frac{\partial}{\partial t} [e^{i(kx - \omega t)}] = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} [e^{i(kx - \omega t)}]$$

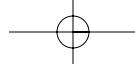
which can be viewed as a reflection of the relation $E = K$ for a free particle.

We have just demonstrated that the equation

$$-i\hbar \frac{\partial}{\partial t} \Psi = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi$$

is correct for the special case of a free particle (potential $U = 0$). Schrödinger generalized this equation by adding a potential energy term, producing his famous result:

$$-i\hbar \frac{\partial}{\partial t} \Psi = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi + U(x) \Psi$$

Section 7.11 • The Time-Dependent Schrödinger Equation **239**

Thus, Schrödinger's time-dependent equation can be viewed as an energy equation: The total energy is kinetic plus potential, $E = K + U$.

Notice that the time-dependent Schrödinger equation involves derivatives with respect to t and x . Since $\Psi(x, t)$ depends on two variables, these derivatives are *partial derivatives*, and the time-dependent Schrödinger equation is a *partial differential equation*.* It has a vast array of solutions, some of which are very complicated, but by far its most important solutions are the stationary states, and these are relatively simple. Recall that we claimed earlier that these are standing waves, with the form (7.11)

$$\Psi(x, t) = \psi(x)e^{-i\omega t} \quad (7.108)$$

and that the spatial wave function $\psi(x)$ had to satisfy the time-*independent* Schrödinger equation. We can now prove these claims.

Let us suppose first that the function (7.108) does satisfy the time-dependent Schrödinger equation (7.107). Substituting into (7.107), we find that

$$i\hbar \frac{\partial}{\partial t} \Psi(x, t) = \hbar\omega \psi(x)e^{-i\omega t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right) \psi(x)e^{-i\omega t}$$

Recognizing that $\hbar\omega = E$ (the de Broglie hypothesis) and canceling the common factor of $e^{-i\omega t}$, we see that

$$E\psi(x) = \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x) \right) \psi(x) \quad (7.109)$$

which is precisely the time-independent Schrödinger equation (7.106). Thus, if the standing wave (7.108) satisfies the time-dependent Schrödinger equation, its spatial part must satisfy the time-independent Schrödinger equation. Conversely, if $\psi(x)$ satisfies the time-independent equation (7.109), we can follow the same steps backward and conclude that the standing wave (7.108) satisfies the time-dependent equation.

Although stationary states, given by the standing waves (7.108), are solutions — and very important solutions — of the time-dependent Schrödinger equation, they are by no means the only solutions. To see this, suppose that we have two stationary states,

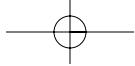
$$\Psi_1(x, t) = \psi_1(x)e^{-i\omega_1 t} \quad \text{and} \quad \Psi_2(x, t) = \psi_2(x)e^{-i\omega_2 t} \quad (7.110)$$

where $\hbar\omega_1 = E_1$ and $\hbar\omega_2 = E_2$ are the corresponding energies. To be definite, we could consider ψ_1 and ψ_2 to be the lowest two wave functions of the rigid box as given in Equation (7.60) with $n = 1$ and 2. Because each of these functions satisfies the time-dependent Schrödinger equation, it is easy to see that the same is true of any linear combination of the form

$$\Psi(x, t) = \alpha\Psi_1(x, t) + \beta\Psi_2(x, t) \quad (7.111)$$

for any two fixed numbers α and β . (That is, the superposition principle applies to the time-dependent Schrödinger equation.) To prove this, just

*There is a brief review of partial derivatives in Section 8.2. If your knowledge of this kind of derivative needs a little refreshing, you might want to read that section before coming back to this.



240 Chapter 7 • The Schrödinger Equation in One Dimension

substitute the proposed solution into the left side of the time-dependent Schrödinger equation:

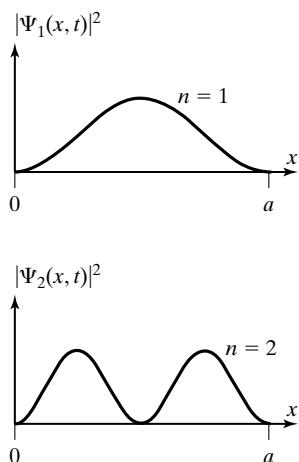


FIGURE 7.24

The probability density $|\Psi(x, t)|^2$ for each of the lowest two energy levels of the rigid box. Because these are stationary states, the probability densities are constant in time.

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \Psi(x, t) &= \alpha i\hbar \frac{\partial}{\partial t} \Psi_1(x, t) + \beta i\hbar \frac{\partial}{\partial t} \Psi_2(x, t) \\ &= \alpha \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right) \Psi_1(x, t) + \beta \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right) \Psi_2(x, t) \\ &= \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right) \Psi(x, t). \end{aligned}$$

The special feature of the separate stationary-state wave functions of (7.110) is that the corresponding probability densities are independent of t (that is, stationary). This is because when we form $|\Psi(x, t)|^2$, the time-dependent exponential factor drops out (since $|e^{-i\omega t}| = 1$). The two probability densities (for the case of the lowest two stationary states of the rigid box) are shown in Fig. 7.24, where each picture is frozen and unchanging in time.

It is perhaps a little surprising that when we form a linear combination of two stationary-state wave functions, as in (7.111), the resulting state is not a stationary state. To see this, we can rewrite (7.111) as

$$\begin{aligned} \Psi(x, t) &= \alpha \psi_1(x) e^{-i\omega_1 t} + \beta \psi_2(x) e^{-i\omega_2 t} \\ &= e^{-i\omega_1 t} [\alpha \psi_1(x) + \beta \psi_2(x) e^{-i\omega_{21} t}] \end{aligned} \quad (7.112)$$

where, in the second line, we have factored out the first exponential $e^{-i\omega_1 t}$ and replaced ω_2 in the second exponential by $\omega_2 - \omega_1$, which we have abbreviated as $\omega_{21} = (\omega_2 - \omega_1)$. When we form the absolute value squared of $\Psi(x, t)$, the first exponential disappears in the now-familiar way, but the second does not.

$$|\Psi(x, t)|^2 = |\alpha \psi_1(x) + \beta \psi_2(x) e^{-i\omega_{21} t}|^2 \quad (7.113)$$

It is fairly easy to see that since one of the terms on the right is oscillatory while the other is not, the right side is definitely not independent of time. The only exception to this statement is if one of the two coefficients α or β is zero, in which case the state is just one of the original stationary states. Rather than prove these statements in general, we examine one particular case in the following example. (But see also Problem 7.59.)

Example 7.6

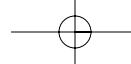
For a particle in a rigid box, consider the nonstationary state with wave function (7.112) for the special case that $\alpha = \beta = 1/\sqrt{2}$. (The particular value $1/\sqrt{2}$ is chosen to guarantee that Ψ is normalized — see Problem 7.58.) Evaluate $|\Psi(x, t)|^2$ and plot it for several different times.

With the given values of α and β , the wave function is

$$\Psi(x, t) = \frac{1}{\sqrt{2}} [\psi_1(x) e^{-i\omega_1 t} + \psi_2(x) e^{-i\omega_2 t}] \quad (7.114)$$

and from (7.113), we find

$$|\Psi(x, t)|^2 = \frac{1}{2} |\psi_1(x) + \psi_2(x) e^{-i\omega_{21} t}|^2 \quad (7.115)$$

Section 7.11 • The Time-Dependent Schrödinger Equation **241**

where the explicit forms of ψ_1 and ψ_2 can be found in (7.60) and $\omega_{21} = (\omega_2 - \omega_1) = (E_2 - E_1)/\hbar$ with the energies given by (7.23). The absolute value on the right can be evaluated in several ways. If you think of it as $|u + ve^{-i\theta}|^2$, then

$$\begin{aligned} |u + ve^{-i\theta}|^2 &= |u + v \cos \theta - iv \sin \theta|^2 \\ &= (u + v \cos \theta)^2 + (v \sin \theta)^2 = u^2 + v^2 + 2uv \cos \theta \end{aligned} \quad (7.116)$$

where in the first line we used the Euler relation for $e^{-i\theta}$ and in the second we used $\cos^2 + \sin^2 = 1$. Thus,

$$|\Psi(x, t)|^2 = \frac{1}{2}[\psi_1(x)^2 + \psi_2(x)^2 + 2\psi_1(x)\psi_2(x) \cos(\omega_{21}t)]. \quad (7.117)$$

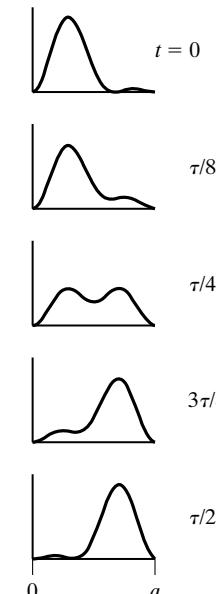
The most obvious thing about this result is that, because of the factor $\cos(\omega_{21}t)$, the probability density *does* vary with time. Notice that in this case the time dependence is actually periodic, with period $\tau = 2\pi/\omega_{21}$.

We have plotted (7.117) for five equally spaced times, $\tau/8$ apart, in Fig. 7.25, where it is quite clear that the probability density is moving around inside the box. At $t = 0$, the probability is concentrated in the left half of the box, with almost no probability of finding the particle on the right. As t advances, the probability shifts across to the right, and by the time $t = \tau/2$, the particle is almost certainly in the right half. If we had shown a few more plots (at the same spacing of t), they would have shown the probability shifting back to the left, and by $t = \tau$ the density would be the same as at $t = 0$ again.

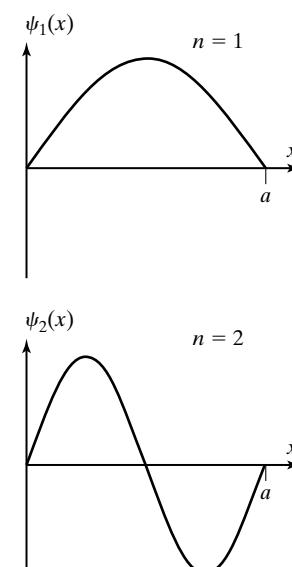
Some features of Fig. 7.25 can be understood if we examine the wave functions carefully. At $t = 0$, the term inside the absolute value signs in (7.115) is just $\psi_1(x) + \psi_2(x)$. If you look at the plots in Fig. 7.26, you will see that $\psi_1(x)$ and $\psi_2(x)$ are both positive on the left of the box but have opposite signs on the right. Therefore, when we add them, they interfere constructively on the left, but destructively on the right. Thus, at $t = 0$, the probability density $|\Psi(x, 0)|^2$ is large on the left and very small on the right, as in the first of the plots of Fig. 7.25. By the time $t = \tau/2$, the exponential factor in (7.115) is $e^{-i\pi} = -1$ and the term inside the absolute value signs is $\psi_1(x) - \psi_2(x)$. Referring again to Fig. 7.26, you will see that now the two waves interfere destructively on the left and constructively on the right, as seen in the last of the plots in Fig. 7.25.

Looking back at (7.114), we can now understand why any linear combination of stationary-state wave functions is itself not stationary: The time dependence of each separate term in (7.114) is contained in a simple exponential factor $e^{-i\omega t}$ (which disappears when we form $|\Psi_1|^2$ or $|\Psi_2|^2$), but these two factors have *different* frequencies ω_1 and ω_2 . Thus, as time goes by, they vary at different rates and the interference between the two terms in the sum keeps changing, as is evident in Fig. 7.25.

The time dependence of nonstationary states, as determined by the time-dependent Schrödinger equation, is fascinating and theoretically very important. Nevertheless — and perhaps fortunately for us — it plays no role in the majority of elementary quantum problems. For the most part, our main concern will be solving the time-independent Schrödinger equation to find the allowed energies and the corresponding stationary-state wave functions for

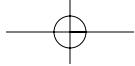
**FIGURE 7.25**

The probability density $|\Psi(x, t)|^2$ for the non-stationary state (7.114) at five equally spaced times.

**FIGURE 7.26**

The two lowest stationary-state wave functions $\psi_1(x)$ and $\psi_2(x)$ for the rigid box.

*If you're not familiar with this result, check it using Euler's formula (7.10).

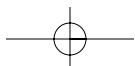


242 Chapter 7 • The Schrödinger Equation in One Dimension

electrons in atoms, molecules, and solids and for nucleons in nuclei. The one exception to this statement concerns the following problem, which we will address in Chapter 11: If the solutions of the time-independent Schrödinger equation were really *perfectly* stationary states, then they would represent states of an atom (or other system) that *never* change. This would imply that an atom in one of its energy levels would never make a transition to another energy level — in flagrant contradiction to the observed transitions made by most atoms. As we will see in Chapter 11, atomic transitions result from various outside influences, such as electromagnetic fields, which mean that the stationary states calculated for the isolated atom are not in fact exactly stationary states for the real system, which must include the outside influences. To calculate how these influences cause the observed transitions, we will need to use the time-dependent Schrödinger equation.

CHECKLIST FOR CHAPTER 7

CONCEPT	DETAILS
Superposition principle	If Ψ_1 and Ψ_2 are possible waves, so is $A\Psi_1 + B\Psi_2$
Classical standing wave Nodes	$\Psi(x, t) = \psi(x) \cos \omega t$ (7.5) (does not move right or left) Points at which the disturbance remains zero for all times
Time dependence of quantum standing waves	$\Psi(x, t) = \psi(x)e^{-i\omega t}$ (7.11) (fixed spatial part, oscillatory time part)
Time-independent Schrödinger equation	$\frac{d^2\psi}{dx^2} = \frac{2m}{\hbar^2} [U(x) - E]\psi \quad (7.39)$
Particle in a one-dimensional rigid box	Moves freely inside $0 \leq x \leq a$ but cannot escape outside
Boundary conditions	$\psi(0) = \psi(a) = 0$ (7.16)
Allowed energies	$E_n = n^2(\pi^2\hbar^2/2ma^2)$ (7.23)
Wave functions	$\psi(x) = \sqrt{2/a} \sin(n\pi x/a)$ (7.60)
Normalization condition	$\int \psi(x) ^2 dx = 1$ (7.55)
Behavior of wave functions	
In classically forbidden zone ($E < U$)	Curves away from axis, $\psi \rightarrow 0$ exponentially as $x \rightarrow \pm\infty$
In classically allowed zone ($E > U$)	Curves toward the axis and oscillates sinusoidally
Allowed energies and well-behaved functions	Section 7.8
Simple harmonic oscillator*	$U = \frac{1}{2}kx^2$ (7.97)
Allowed energies	$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_c$ (7.98)
Tunneling*	Ability of a quantum particle to penetrate into a classically forbidden zone
Probability of tunneling	$P = e^{-2\alpha L}$ (7.105)
Time-dependent Schrödinger equation*	$i\hbar \frac{\partial}{\partial t} \Psi(x, t) = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x)\right) \Psi(x, t)$ (7.107)



PROBLEMS FOR CHAPTER 7

SECTION 7.2 (Classical Standing Waves)

- 7.1** • A string is oscillating with wave function

$$y(x, t) = A \sin kx \cos \omega t \quad (7.118)$$

with $A = 3 \text{ cm}$, $k = 0.2\pi \text{ rad/cm}$, and $\omega = 10\pi \text{ rad/s}$. For each of the times $t = 0, 0.05$, and 0.07 s , sketch the string for $0 \leq x \leq 10 \text{ cm}$.

- 7.2** • A standing wave of the form (7.118) has amplitude 2 m, wavelength 10 m, and period $T = 2 \text{ s}$. Write down the expression for $y(x, t)$.
- 7.3** • A string is clamped at the points $x = 0$ and $x = 30 \text{ cm}$. It is oscillating sinusoidally with amplitude 2 cm, wavelength 20 cm, and frequency $f = 40 \text{ Hz}$. Write an expression for its wave function $y(x, t)$.
- 7.4** • For the standing wave of Eq. (7.118) calculate the string's transverse velocity at any fixed position x . [That is, differentiate (7.118) with respect to t , treating x as a constant.] At certain times the string's displacement is zero for all x , and an instantaneous snapshot of the string would show no wave at all. Sketch the string at such a moment, and indicate, with several small arrows, the velocity of several points on the string at that moment.
- 7.5** • In Fig. 7.27 are sketched three successive snapshots of a standing wave on a string. The first (solid) curve shows the string at maximum displacement. Were these snapshots taken at equally spaced times? If not, make a sketch in which they were. (Use the same first and third curves.)



FIGURE 7.27

(Problems 7.5 and 7.6)

- 7.6** • Figure 7.27 shows a standing wave on a string. In any position other than the dotted one, the string has potential energy because it is stretched (compared to the straight, dotted position). Let U_0 denote this potential energy at the position of the solid curve (which shows the maximum displacement). What is the potential energy when the string reaches the position of the dotted line? What is the kinetic energy at this position?

- 7.7** • Consider a standing wave on a string, clamped at the points $x = 0$ and $x = 40 \text{ cm}$. It is oscillating with amplitude 3 cm and wavelength 80 cm, and its maximum transverse velocity is 60 cm/s. Write an expression for its displacement $y(x, t)$ as a function of x and t .

- 7.8** • Rewrite the expression (7.118) for a standing wave, eliminating the wave number k and angular frequency ω in favor of the wavelength λ and period T . Use

your answer to explain clearly why the wave repeats itself if we move from any point x_0 to $x_0 + \lambda$ (with t fixed). Show similarly what happens if we increase t from any t_0 to $t_0 + T$.

- 7.9** • A string is oscillating with wave function of the form (7.118) but with $A = 2.5 \text{ cm}$, $k = 1 \text{ rad/cm}$, and $\omega = 10\pi \text{ rad/s}$. **(a)** Sketch two complete wavelengths of the wave at each of the times $t = 0.05 \text{ s}$ and $t = 0.1 \text{ s}$. **(b)** For a fixed value of x , differentiate (7.118) with respect to t to give the string's transverse velocity at any position x . Graph this velocity as a function of x for each of the two times in part (a).

SECTION 7.3 (Standing Waves in Quantum Mechanics)

- 7.10** • Prove that any complex number $z = x + iy$ (with x and y real) can be written as $z = re^{i\theta}$ (with r and θ real). Give expressions for x and y in terms of r and θ , and vice versa. [Hint: Use Euler's formula, (7.10).]

- 7.11** • **(a)** Show for any complex number $z = x + iy$, that $|z|^2 = zz^*$ where $z^* = x - iy$ is the complex conjugate of z . **(b)** Hence, prove the useful identity, $|zw| = |z| \cdot |w|$. **(c)** Show for any standing wave, with the form (7.11), $|\Psi(x, t)| = |\psi(x)|$, and hence that the probability density is independent of time.

- 7.12** • Consider a complex function of time, $z = re^{-i\omega t}$, where r is a real constant. **(a)** Write z in terms of its real and imaginary parts, x , and y , and show that they oscillate sinusoidally and 90° out of phase. **(b)** Show that $|z| = \sqrt{x^2 + y^2}$ is constant.

- 7.13** • We claimed in connection with Eq. (7.7) that the general (real) sinusoidal wave has time dependence

$$a \cos \omega t + b \sin \omega t \quad (7.119)$$

Another way to say this is that the general sinusoidal time dependence is

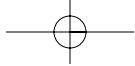
$$A \sin(\omega t + \phi) \quad (7.120)$$

- (a)** Show that these two forms are equivalent; that is, prove that the function (7.119) can be expressed in the form (7.120) and vice versa. Give expressions for a and b in terms of A and ϕ and vice versa. (Remember the trig identities in Appendix B.) **(b)** Show that by changing the origin of time (that is, rewriting everything in terms of $t' = t + \text{constant}$, with a suitably chosen constant) you can eliminate the constant ϕ from (7.120) [that is, rewrite (7.120) as $A \sin \omega t'$.]

- 7.14** • The function e^z is defined for any z , real or complex, by its power series

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$$

Write down this series for the case that z is purely imaginary, $z = i\theta$. Note that the terms in this series are alternately real and imaginary. Group together all



244 Chapter 7 • The Schrödinger Equation in One Dimension

the real terms and all the imaginary terms, and prove the important identity, called Euler's formula,

$$e^{i\theta} = \cos \theta + i \sin \theta \quad (7.121)$$

[Hint: You will need to remember the power series for $\cos \theta$ and $\sin \theta$ given in Appendix B. You may wonder whether it is legitimate to regroup the terms of an infinite series as recommended here; for power series like those in this problem, it is legitimate.]

- 7.15** •• Use the identity $e^{i(\theta+\phi)} = e^{i\theta}e^{i\phi}$ to prove the trig identities for $\cos(\theta + \phi)$ and $\sin(\theta + \phi)$.

- 7.16** •• Use the result (7.121) to prove that

$$\cos \theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \quad \text{and} \quad \sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2}$$

SECTION 7.4 (The Particle in a Rigid Box)

- 7.17** • Find the lowest three energies, in eV, for an electron in a one-dimensional box of length $a = 0.2$ nm (about the size of an atom).

- 7.18** • Find the lowest three energies, in MeV, of a proton in a one-dimensional rigid box of length $a = 5$ fm (a typical nuclear size).

- 7.19** • What is the spacing, in eV, between the lowest two levels of an electron confined in a one-dimensional wire of length 1 cm?

- 7.20** • Sketch the energy levels and wave functions for the levels $n = 5, 6, 7$ for a particle in a one-dimensional rigid box. (See Fig. 7.5.)

- 7.21** • For the ground state of a particle in a rigid box, we have seen that the momentum has a definite magnitude $\hbar k$ but is equally likely to be in either direction. This means that the uncertainty in p is $\Delta p \approx \hbar k$. The uncertainty in position is $\Delta x \approx a/2$. Verify that these uncertainties are consistent with the Heisenberg uncertainty principle (6.34).

SECTION 7.6 (The Rigid Box Again)

- 7.22** • Prove that the function $\psi = Ae^{\alpha x} + Be^{-\alpha x}$ satisfies the equation $\psi'' = \alpha^2\psi$ for any two constants A and B .

- 7.23** • We saw that the coefficients A and B in the wave function for a negative-energy state in the rigid box would have to satisfy $A + B = 0$ and $Ae^{\alpha a} + Be^{-\alpha a} = 0$. Show that this is possible only if $A = B = 0$ and hence, that there are no negative-energy states.

- 7.24** • Prove that the function $\psi = Ae^{ikx} + Be^{-ikx}$ satisfies the equation $\psi'' = -k^2\psi$ for any two constants A and B .

- 7.25** •• We have seen that second-order differential equations like the Schrödinger equation have two independent solutions. Consider the fourth-order equation $d^4\psi/dx^4 = \beta^4\psi$, where β is a positive constant. Prove that each of the four functions $e^{\pm\beta x}$ and $e^{\pm i\beta x}$ is a solution. (It is a fact — though harder to prove — that any solution of the equation can be expressed as a linear combination of these four solutions.)

- 7.26** •• A second-order differential equation like the Schrödinger equation has two independent solutions $\psi_1(x)$ and $\psi_2(x)$. These two solutions can be chosen in many ways, but once they are chosen, any solution can be expressed as a linear combination $A\psi_1(x) + B\psi_2(x)$ (where A and B are constants, real or complex). (a) To illustrate this property, consider the differential equation $\psi'' = -k^2\psi$ where k is a constant. Prove that each of the three functions $\sin kx$, $\cos kx$, and e^{ikx} is a solution. (b) Show that each can be expressed as a combination of the other two.

- 7.27** •• Many physical problems lead to a differential equation of the form

$$a\psi''(x) + b\psi'(x) + c\psi(x) = 0$$

where a, b, c are constants. [An example was given in (7.30).] (a) Prove that this equation has two solutions of the form $\psi(x) = e^{\gamma x}$, where γ is either solution of the quadratic equation $ay^2 + by + c = 0$. (b) Prove that any linear combination of these two solutions is itself a solution.

- 7.28** •• Consider the second-order differential equation

$$f(x)\psi''(x) + g(x)\psi'(x) + h(x)\psi(x) = 0$$

where f, g , and h are known functions of x . Prove that if $\psi_1(x)$ and $\psi_2(x)$ are both solutions of this equation, the linear combination $A\psi_1(x) + B\psi_2(x)$ is also a solution for any two constants A and B — the result known as the superposition principle.

- 7.29** •• Show that the integral which appears in the normalization condition for a particle in a rigid box has the value

$$\int_0^a \sin^2\left(\frac{n\pi x}{a}\right) dx = \frac{a}{2}$$

[Hint: Use the identity for $\sin^2 \theta$ in terms of $\cos 2\theta$ given in Appendix B.]

- 7.30** •• (a) Write down and sketch the probability distribution $|\psi(x)|^2$ for the second excited state ($n = 3$) of a particle in a rigid box of length a . (b) What are the most probable positions, x_{mp} ? (c) What are the probabilities of finding the particle in the intervals $[0.50a, 0.51a]$ and $[0.75a, 0.76a]$?

- 7.31** •• Answer the same questions as in Problem 7.30 but for the third excited state ($n = 4$).

- 7.32** •• If a particle has wave function $\psi(x)$, the probability of finding the particle between any two points b and c is

$$P(b \leq x \leq c) = \int_b^c |\psi(x)|^2 dx \quad (7.122)$$

For a particle in the ground state of a rigid box, calculate the probability of finding it between $x = 0$ and $x = a/3$ (where a is the width of the box). Use the hint in Problem 7.29.

- 7.33** •• Evaluate the integral (7.65) to give the average result $\langle x \rangle$ (the expectation value) found when the position of a particle in the ground state of a rigid box is measured many times. [Hint: Rewrite $\sin^2(\pi x/a)$ in terms of $\cos(2\pi x/a)$ and use integration by parts.]
- 7.34** •• Do Problem 7.33 for the case of the second excited state ($n = 3$) of a rigid box.
- 7.35** ••• Consider a particle in the ground state of a rigid box of length a . (a) Evaluate the integral (7.122) to give the probability of finding the particle between $x = 0$ and $x = c$ for any $c \leq a$. (b) What does your result give when $c = a$? Explain. (c) What if $c = a/2$? (d) What if $c = a/4$? (e) The answer to part (c) is half that for part (b), whereas that to part (d) is *not* half that for part (c). Explain.

SECTION 7.8 (The Nonrigid Box)

- 7.36** • Consider the potential-energy function $U(x) = \frac{1}{2}kx^2$. (a) Sketch U as a function of x . (b) For a classical particle of energy E , find the turning points in terms of E and k . (c) If we double the energy, what happens to the length of the classically allowed region?
- 7.37** •• Consider the potential-energy function (the Gaussian well)

$$U(x) = U_0(1 - e^{-x^2/a^2})$$

where U_0 and a are positive constants. (a) Sketch $U(x)$. (b) For $0 < E < U_0$, find the classical turning points (in terms of U_0 , E , and a).

- 7.38** •• Give an argument that a particle moving in either of the finite wells of Figs. 7.8(b) and (c) can have no states with $E < 0$. [Hint: Remember that whenever $U(x) - E$ is positive, $\psi(x)$ must bend away from the axis; show that a wave function that is well behaved as $x \rightarrow -\infty$ (that is, behaves like $e^{\alpha x}$) necessarily blows up as $x \rightarrow \infty$.]

- 7.39** •• Make sketches of the probability density $|\psi(x)|^2$ for the first four energy levels of a particle in a non-rigid box. Draw next to each graph the corresponding graph for the rigid box.

- 7.40** •• Consider the potential well shown in Fig. 7.28(a). Sketch the wave function for the eighth excited level of this well, assuming that it has energy $E > U_0$. [Hint: The wave function has eight nodes — not counting the nodes at the walls of the well.]

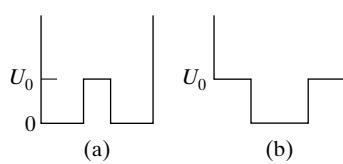


FIGURE 7.28
(Problems 7.40 to 7.43)

- 7.41** •• Do the same task as in Problem 7.40, but for the well in Fig. 7.28(b).

- 7.42** •• Sketch the wave function for the fourth excited level of the well of Fig. 7.28(a), assuming that it has energy in the interval $0 < E < U_0$.

- 7.43** •• Do Problem 7.42, but for the well of Fig. 7.28(b).

- 7.44** •• Consider the infinitely deep potential well shown in Fig. 7.29. (a) Argue that this is the potential energy of a particle of mass m above a hard surface in a uniform gravitational field with x measured vertically up. (b) Sketch the wave functions for the ground state and the first two excited states of this well.

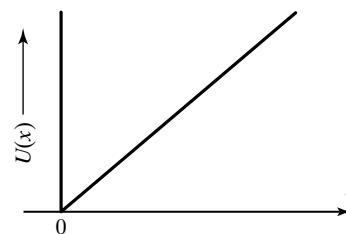


FIGURE 7.29

(Problem 7.44)

- 7.45** •• (a) Derive an approximate expression for the number of bound states in a finite square well of depth U_0 . To do this, make the assumption that the energy of the highest state of the finite well is close to the corresponding energy of the infinite well. This state is the one that has an energy E near the top of the well, that is, $E = U_0$. (b) Compute the approximate number of bound states for the case of an electron in a well of depth $U_0 = 10$ eV, and width $a = 0.3$ nm. These numbers crudely approximate valence electrons in a solid.

- 7.46** •• In a region where the potential energy $U(x)$ varies, the Schrödinger equation must usually be solved by numerical methods — generally with the help of a computer. The simplest such method, though not the most efficient, is called Euler's method and can be described as follows: In solving the Schrödinger equation numerically, one needs to know the values of $\psi(x)$ and $\psi'(x)$ at one point x_0 . For example, for the finite wells of Fig. 7.8 we know these values at $x_0 = 0$ since we know that $\psi(x)$ has the form $\psi(x) = e^{ax}$ for $x \leq 0$. Suppose now we want to find $\psi(x)$ at some point $x > 0$. We first divide the interval from x_0 to x into n equal intervals each of width Δx :

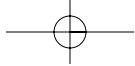
$$x_0 < x_1 < x_2 < \dots < x_n = x$$

Knowing $\psi(x)$ and $\psi'(x)$ at $x = x_0$, we can use the Schrödinger equation to find approximate values of these two functions at x_1 , and from these, we can find values at x_2 and so on until we know both functions at $x_n = x$. To do this, we use the approximations

$$\psi(x_{i+1}) \approx \psi(x_i) + \psi'(x_i) \Delta x$$

and

$$\psi'(x_{i+1}) \approx \psi'(x_i) + \psi''(x_i) \Delta x$$



246 Chapter 7 • The Schrödinger Equation in One Dimension

[The first approximation is just the definition of the derivative, and the second is the same approximation applied to $\psi'(x)$.] Consider the equation $\psi''(x) = -\psi(x)$, with the starting values $\psi(0) = 0$ and $\psi'(0) = 1$, and do the following (for which you don't need a computer): (a) What is the exact solution of this equation with these initial conditions? What is the exact value of $\psi(x)$ at $x = 1$? (b) Divide the interval from $x = 0$ to 1 into two equal intervals ($n = 2$), and use the Euler's method to find an approximate value for $\psi(1)$. (c) Repeat with $n = 3$ and $n = 4$. Compare your results with the exact answer, and make a plot of the exact and approximate values for the case that $n = 4$. Note well how the approximate solution improves as you increase n .

- 7.47** •• Do Problem 7.46, but for the differential equation $\psi''(x) = +\psi(x)$, with the starting values $\psi(0) = 0$ and $\psi'(0) = 1$.

SECTION 7.9 (The Simple Harmonic Oscillator)

- 7.48** • Show that the length parameter b defined for the SHO in Eq. (7.99) is equal to the value of x at the classical turning point for a particle with the energy of the quantum ground state.

- 7.49** • Verify that the $n = 0$ wave function for the SHO, given in Table 7.1, satisfies the Schrödinger equation with $E = \frac{1}{2}\hbar\omega_c$. (The potential-energy function is $U = \frac{1}{2}kx^2$.)

- 7.50** • The wave function $\psi_0(x)$ for the ground state of a harmonic oscillator is given in Table 7.1. Show that its normalization constant A_0 is

$$A_0 = (\pi b^2)^{-1/4} \quad (7.123)$$

You will need to know the integral $\int_{-\infty}^{\infty} e^{-\lambda x^2} dx$, which can be found in Appendix B.

- 7.51** • Show that the normalization constant A_1 for the wave function of the first excited state of the SHO is $A_1 = (4/\pi b^2)^{1/4}$. (The wave function is given in Table 7.1. You will need to know the integral $\int_{-\infty}^{\infty} x^2 e^{-\lambda x^2} dx$, which is given in Appendix B.)

- 7.52** •• Verify that the $n = 2$ wave function for the SHO, given in Table 7.1, satisfies the Schrödinger equation with energy $\frac{5}{2}\hbar\omega_c$.

- 7.53** ••• The wave functions of the harmonic oscillator, like those of a particle in a finite well, are nonzero in the classically forbidden regions, outside the classical turning points. In this question you will find the probability that a quantum particle which is in the ground state of an SHO will be found outside its classical turning points. The wave function for this state is in Table 7.1, and its normalization constant A_0 is given in Problem 7.50. (a) What are the turning points for a classical particle with the ground-state energy $\frac{1}{2}\hbar\omega_c$ in an SHO with $U = \frac{1}{2}kx^2$? Relate your answer to the constant b in Eq. (7.99). (b) For a quantum particle in the ground state, write down the integral that gives the total probability for finding the particle between the two classical turning points. The form of the required integral is given in Problem 7.32, Eq. (7.122).

To evaluate it, change variables until you get an integral of the form $\int_{-1}^1 e^{-y^2} dy$; this is a standard integral of mathematical physics (called the error function) with the known value 1.49. What is the probability of finding the particle between the classical turning points? (c) What is the probability of finding it *outside* the classical turning points?

SECTION 7.10 (Tunneling)

- 7.54** • Consider two straight wires lying on the x axis, separated by a gap of 4 nanometers. The potential energy U_0 in the gap is about 3 eV higher than the energy of a conduction electron in either wire. What is the probability that a conduction electron in one wire arriving at the gap will pass through the gap into the other wire?

- 7.55** •• The radioactive decay of certain heavy nuclei by emission of an alpha particle is a result of quantum tunneling, as described in detail in Section 17.10. Meanwhile, here is a simplified model: Imagine an alpha particle moving around inside a nucleus, such as thorium 232. When the alpha bounces against the surface of the nucleus, it meets a barrier caused by the attractive nuclear force. The dimensions of this barrier vary a lot from one nucleus to another, but as representative numbers you can assume that the barrier's width is $L \approx 35$ fm ($1 \text{ fm} = 10^{-15} \text{ m}$) and the average barrier height is $U_0 - E \approx 5 \text{ MeV}$. Find the probability that an alpha hitting the nuclear surface will escape. Given that the alpha hits the nuclear surface about 5×10^{21} times per second, what is the probability that it will escape in a day?

SECTION 7.11 (The Time-Dependent Schrödinger Equation)

- 7.56** • Verify that $e^{i\pi} = -1$.

- 7.57** • If $\psi(x)$ satisfies the time-independent Schrödinger equation (7.106), verify that the function $\Psi(x, t) = \psi(x)e^{-iEt/\hbar}$ satisfies the time-dependent Schrödinger equation (7.107).

- 7.58** •• Assuming the wave functions ψ_1 and ψ_2 are normalized, verify that the superposition (7.114) is also normalized. [Hint: It will help if you can prove that $\int \psi_1 \psi_2 dx = 0$.]

- 7.59** •• Verify that for the wave function (7.111),

$$|\Psi(x, t)|^2 = |\alpha|^2 |\psi_1(x)|^2 + |\beta|^2 |\psi_2(x)|^2 + 2\text{Re}(\alpha^* \beta) \psi_1(x) \psi_2(x) \cos(\omega_{21}t)$$

assuming that $\psi_1(x)$ and $\psi_2(x)$ are real.

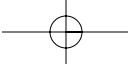
COMPUTER PROBLEMS

- 7.60** • (Section 7.2) Use appropriate software to draw five graphs (all on the same plot) of the standing wave (7.2) for $0 < x < 2$ and for the five times $t = 0, 0.05, 0.1, 0.15$, and 0.2 . Take $A = \lambda = T = 1$. Using your plots, describe clearly the behavior of the standing wave.

- 7.61** • (Section 7.9) Use appropriate software to plot the lowest two wave functions of the simple harmonic oscillator. The functions are given in Table 7.1, and the normalization constants in Problems 7.50 and 7.51. For the purposes of your plot, you may as well choose your unit of length such that $b = 1$.
- 7.62** •• (Section 7.2) If you have access to graphing software that lets you animate graphics, make 20 separate plots of the standing wave of Problem 7.60 at equally spaced times $t = 0, 0.05, 0.01, \dots, 0.95$ and animate them to make a movie of the standing wave. Describe the behavior of the wave.
- 7.63** •• (Section 7.11) Make plots similar to Fig. 7.25, but for eight different times, $t = 0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \dots, \frac{7}{8}$. If possible, animate these pictures to show the movement of the distribution in the well.
- 7.64** ••• (Section 7.8) If you haven't already done so, do Problem 7.46, then using a programmable calculator or computer software such as MathCad, extend your answers to the cases that the number of steps n is 5, 10, and 50. Make a plot showing the exact solution and the values of your approximate solution for $n = 50$.
- 7.65** ••• (Section 7.8) If you haven't already done so, do Problem 7.47, then using a programmable calculator or computer software such as MathCad, extend your answers to the cases that the number of steps n is 5, 10, and 50. Make a plot showing the exact solution and the values of your approximate solution for $n = 50$.
- 7.66** ••• (Section 7.8) If you have access to computer software with preprogrammed numerical solution of differential equations (for example, the function NDSolve in Mathematica), do the following: **(a)** Plot the Gaussian well of Problem 7.37, using units such that $U_0 = a = \hbar = 1$ and taking $m = 36$. (The first three choices simply define a convenient system of units — for instance, a is the unit of length — and the last is chosen so that the well supports several bound states.) **(b)** We know that to be acceptable, the wave

function must be proportional to $e^{\alpha x}$ far to the left of the well. To ensure this, use the boundary conditions $\psi(-3) = 1$ and $\psi'(-3) = \alpha$, and solve the Schrödinger differential equation for energy $E = 0.1$. Plot your solution, and confirm that it does not behave acceptably as $x \rightarrow \infty$. **(c)** Repeat for $E = 0.2$. What can you conclude about the energy of the ground state from the plots of parts (b) and (c)? **(d)** Repeat for two or three intermediate energies until you know the ground-state energy to two significant figures.

- 7.67** ••• (Section 7.8) Consider the problem of a particle in finite square well of depth U_0 , as described in Section 7.8. To solve this problem, use a coordinate system centered on the box, with the left edge of the well at $x = -a/2$ and the right edge at $x = +a/2$. From the symmetry of the potential, one can argue that the stationary states should have symmetric probability distributions, that is, $|\psi(x)|^2 = |\psi(-x)|^2$. This implies that solutions are either *symmetric* and satisfy $\psi(x) = \psi(-x)$ or are *antisymmetric* and satisfy $\psi(x) = -\psi(-x)$. Thus, with our choice of coordinates, the solutions within the well are either of the form $\cos(kx)$ (for $n = 1, 3, 5, \dots$) or of the form $\sin(kx)$ (for $n = 2, 4, 6, \dots$). **(a)** For the case of the symmetric solutions, derive an equation relating k and α . [Hint: Using the boundary conditions that both $\psi(x)$ and $\psi'(x)$ are continuous at $x = -a/2$, you can produce two equations that relate k , α , and the arbitrary coefficients A and G in Section 7.8. By dividing these equations, you can eliminate the coefficients. The final equation you produce is a *transcendental equation* relating E and U_0 ; it cannot be solved for E using elementary methods.] **(b)** Show that the equation derived in (a) produces the correct stationary-state energies in the limit $U_0 \rightarrow \infty$. **(c)** Using numerical techniques, find the ground-state energy E_1 of a square well of depth $U_0 = 3 E_1$. [Hint: A simple trial-and-error search for the solution works surprisingly well.]



Chapter 8

The Three-Dimensional Schrödinger Equation

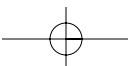
- 8.1** Introduction
 - 8.2** The Three-Dimensional Schrödinger Equation and Partial Derivatives
 - 8.3** The Two-Dimensional Square Box
 - 8.4** The Two-Dimensional Central-Force Problem
 - 8.5** The Three-Dimensional Central-Force Problem
 - 8.6** Quantization of Angular Momentum
 - 8.7** The Energy Levels of the Hydrogen Atom
 - 8.8** Hydrogenic Wave Functions
 - 8.9** Shells
 - 8.10** Hydrogen-Like Ions
- Problems for Chapter 8

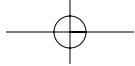
8.1 Introduction

In Chapter 7 we studied the one-dimensional Schrödinger equation and saw how it determines the allowed energies and corresponding wave functions of a particle in one dimension. If the world in which we lived were one-dimensional, we could now proceed to apply these ideas to various real systems: atoms, molecules, nuclei, and so on. However, our world is three-dimensional, and we must first describe how the one-dimensional equation is generalized to three dimensions.

We will find that the three-dimensional equation is appreciably more complicated than its one-dimensional counterpart, involving derivatives with respect to all three coordinates x , y , and z . Nevertheless, its most important properties will be familiar. Specifically, in three dimensions, just as in one dimension (and two), the time-independent Schrödinger equation is a differential equation for the wave function ψ . For most systems, this equation has acceptable solutions only for certain particular values of the energy E . Those E for which it has an acceptable solution are the allowed energies of the system, and the solutions ψ are the corresponding wave functions.

In this chapter we write down the three-dimensional Schrödinger equation and describe its solutions for some simple systems, culminating with the hydrogen atom. Since many of the important features of the three-dimensional equation are already present in the simpler case of two dimensions, two of our examples will be two-dimensional.





8.2 The Three-Dimensional Schrödinger Equation and Partial Derivatives

In one dimension the wave function $\psi(x)$ for a particle depends on the one coordinate x , and the time-independent Schrödinger equation has the now familiar form

$$\frac{d^2\psi}{dx^2} = \frac{2M}{\hbar^2}[U - E]\psi \quad (8.1)$$

where U is the particle's potential energy and we temporarily use capital M for the particle's mass.* In this equation, remember that both ψ and U are functions of x , whereas E , although it can take on various values, does not depend on x . In three dimensions we would naturally expect the wave function ψ to depend on all three coordinates x , y , and z ; that is,

$$\psi = \psi(x, y, z) = \psi(\mathbf{r})$$

where

$$\mathbf{r} = (x, y, z)$$

Similarly the potential energy U will normally depend on x , y , and z :

$$U = U(x, y, z) = U(\mathbf{r})$$

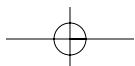
How the differential equation (8.1) generalizes to three dimensions is not so obvious. Here we will simply state that the correct generalization of (8.1) is this:

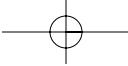
$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{2M}{\hbar^2}[U - E]\psi \quad (8.2)$$

where the three derivatives on the left are the so-called *partial derivatives*, whose definition and properties we discuss in a moment. That (8.2) is a *possible* generalization of (8.1) is perhaps fairly obvious. That it is the *correct* generalization is certainly not obvious. In more advanced texts you will find various arguments that *suggest* the correctness of (8.2). However, the ultimate test of any equation is whether its predictions agree with experiment, and we will see that the three-dimensional Schrödinger equation in the form (8.2) has passed this test repeatedly when applied to atomic and subatomic systems.

The three derivatives that appear in (8.2) are called **partial derivatives**, and it is important that you understand what these are. An *ordinary* derivative, like $d\psi/dx$, is defined for a function such as $\psi(x)$ that depends on just one variable (for example, the temperature as a function of position x along a narrow rod). Partial derivatives arise when one considers functions of two or

*In the context of the three-dimensional Schrödinger equation, the letter m is traditionally used for the integer that labels the allowed values of the components of angular momentum. It is to avoid confusion with this notation that we use M for the mass in the first six sections of this chapter. In Section 8.7 we return to the hydrogen atom, in which the relevant particle is the electron, whose mass we will call m_e .





250 Chapter 8 • The Three-Dimensional Schrödinger Equation

more variables, such as the temperature as a function of position (x, y, z) in a three-dimensional room. If ψ depends on three variables x, y, z , we define the partial derivative $\partial\psi/\partial x$ as the derivative of ψ with respect to x , obtained when we hold y and z fixed. Similarly, $\partial\psi/\partial y$ is the derivative with respect to y , when x and z are held constant; and similarly with $\partial\psi/\partial z$. Notice that it is customary to use the symbol ∂ for this new kind of derivative.

The calculation of partial derivatives is very simple in practice, as the following example shows.

Example 8.1

Find the three partial derivatives $\partial\psi/\partial x$, $\partial\psi/\partial y$, and $\partial\psi/\partial z$ for $\psi(x, y, z) = x^2 + 2y^3z + z$.

If y and z are held constant, then ψ has the form

$$\psi = x^2 + \text{constant}$$

and the rules of ordinary differentiation give

$$\frac{\partial\psi}{\partial x} = 2x$$

If instead we hold x and z constant, then ψ has the form

$$\psi = \text{constant} + 2zy^3$$

Since the coefficient $2z$ is constant, the rules of ordinary differentiation give

$$\frac{\partial\psi}{\partial y} = 6zy^2$$

Finally, if x and y are constant, then

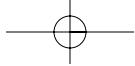
$$\psi = \text{constant} + 2y^3z + z$$

and since $2y^3$ is constant,

$$\frac{\partial\psi}{\partial z} = 2y^3 + 1$$

Higher partial derivatives are defined similarly. For example, $\partial^2\psi/\partial x^2$ is the second derivative of ψ with respect to x , obtained if we hold y and z constant, and so on. Partial derivatives are really no harder to use than ordinary derivatives, once you understand their definition. If you have never worked with them before, you might find it useful to try some of Problems 8.1 to 8.7 right away.

An equation, like the Schrödinger equation (8.2), that involves partial derivatives is called a *partial differential equation*. Now that we know what partial derivatives are, we can discuss methods for solving such equations, and this is what we will do in the remainder of this chapter.



8.3 The Two-Dimensional Square Box

Before we consider any three-dimensional systems, we consider an example in two dimensions. This shares several important features of three-dimensional systems, but is naturally somewhat simpler.

Looking at the Schrödinger equations (8.1) and (8.2) in one and three dimensions, it is easy to guess, correctly, that the Schrödinger equation for a particle of mass M in two dimensions should read

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = \frac{2M}{\hbar^2}[U - E]\psi \quad (8.3)$$

Here ψ is a function of the two-dimensional coordinates, $\psi = \psi(\mathbf{r}) = \psi(x, y)$, and $U = U(\mathbf{r})$ is the particle's potential energy.

The method for solving the Schrödinger equation depends on the potential-energy function $U(\mathbf{r})$. In this section we consider a particle confined in a two-dimensional rigid square box or square well, that is, a particle for which $U(\mathbf{r})$ is zero inside a square region like that shown in Fig. 8.1, but is infinite outside.

$$U(x, y) = \begin{cases} 0 & \text{when } 0 \leq x \leq a \text{ and } 0 \leq y \leq a \\ \infty & \text{otherwise} \end{cases} \quad (8.4)$$

A classical example of such a system would be a metal puck sliding on a frictionless square air table with perfectly rigid, elastic bumpers at its edges. As a quantum example, we could imagine an electron confined inside a thin square metal sheet.

A classical particle inside a rigid square box would bounce indefinitely inside the box. Since $U = 0$ inside the box, its energy E would be all kinetic and could have any value in the range

$$0 \leq E < \infty.$$

To find the possible energies for the corresponding quantum system, we must solve the Schrödinger equation (8.3) with the potential-energy function (8.4).

Since the particle cannot escape from the box, the wave function $\psi(x, y)$ is zero outside the box, and since $\psi(x, y)$ must be continuous, it must also be zero on the boundary:

$$\psi(x, y) = 0, \quad \text{if } x = 0 \text{ or } a, \text{ and if } y = 0 \text{ or } a \quad (8.5)$$

Since $U(x, y) = 0$ inside the box, the Schrödinger equation reduces to

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = -\frac{2ME}{\hbar^2}\psi \quad (8.6)$$

for all x and y inside the box. We must solve this equation, subject to the boundary conditions (8.5).

Separation of Variables

If we knew nothing at all about partial differential equations, the solution of (8.6) would be a formidable prospect. Fortunately, there is an extensive mathematical theory of partial differential equations, which tells us that equations

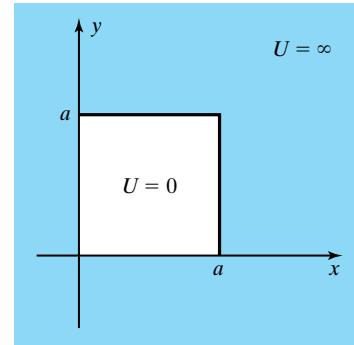
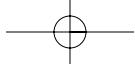


FIGURE 8.1

The two-dimensional, rigid square box. The particle is confined by perfectly rigid walls to the unshaded, square region, within which it moves freely.



252 Chapter 8 • The Three-Dimensional Schrödinger Equation

like (8.6) can be solved by a method called **separation of variables**. In this method one seeks solutions with the form

$$\psi(x, y) = X(x)Y(y) \quad (8.7)$$

where $X(x)$ is a function of x alone and $Y(y)$ a function of y alone. We describe a function with the form (8.7) as a *separated function*.* It is certainly not obvious that there will be any solutions with this separated form. On the other hand, there is nothing to stop us from seeing if there are such solutions, and we will find that indeed there are. Furthermore, the mathematical theory of equations like (8.6) guarantees that *any* solution of the equation can be expressed as a sum of separated solutions. This means that once we have found all of the solutions with the form (8.7), we have, in effect, found *all* solutions. It is this mathematical theorem that is the ultimate justification for using separation of variables to solve the Schrödinger equation for many two- and three-dimensional systems.

To see whether the Schrödinger equation (8.6) does have separated solutions with the form (8.7), we substitute (8.7) into (8.6). When we do this, two simplifications occur. First, the partial derivatives simplify. Consider, for example,

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{\partial^2}{\partial x^2}[X(x)Y(y)]$$

This derivative is evaluated by treating y as fixed. Therefore, the term $Y(y)$ can be brought outside, to give

$$\frac{\partial^2 \psi}{\partial x^2} = Y(y) \frac{\partial^2}{\partial x^2} X(x)$$

Since $X(x)$ depends only on x , the remaining derivative is an ordinary derivative, which we can write as

$$\frac{\partial^2}{\partial x^2} X(x) = \frac{d^2}{dx^2} X(x) = X''(x)$$

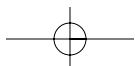
where, as usual, the double prime indicates the second derivative of the function concerned. Thus

$$\frac{\partial^2 \psi}{\partial x^2} = Y(y)X''(x) \quad (8.8)$$

Similarly,

$$\frac{\partial^2 \psi}{\partial y^2} = X(x)Y''(y) \quad (8.9)$$

*Note that by no means can every function of x and y be separated in this way. As simple a function as $x + y$ cannot be expressed as the product of one function of x and one of y .



Section 8.3 • The Two-Dimensional Square Box **253**

If we substitute (8.8) and (8.9) into the Schrödinger equation (8.6), we find that

$$Y(y)X''(x) + X(x)Y''(y) = -\frac{2ME}{\hbar^2}X(x)Y(y)$$

To separate the terms that depend on x from those that depend on y , we divide by $X(x)Y(y)$ to give

$$\frac{X''(x)}{X(x)} + \frac{Y''(y)}{Y(y)} = -\frac{2ME}{\hbar^2} \quad (8.10)$$

The right side of this equation is constant (independent of x and y). Thus (8.10) has the general form

$$(\text{function of } x) + (\text{function of } y) = \text{constant}$$

for all x and y (in the box). To see what this implies, we move the function of y over to the right:

$$(\text{function of } x) = \text{constant} - (\text{function of } y).$$

This equation asserts that a certain function of x is equal to a quantity that does not depend on x at all. In other words, this function, which can depend only on x , is in fact independent of x . This is possible only if the function in question is a constant. We conclude that the quantity $X''(x)/X(x)$ in (8.10) is a constant:

$$\frac{X''(x)}{X(x)} = \text{constant} \quad (8.11)$$

If we call this constant $-k_x^2$, (8.11) can be rewritten as

$$X''(x) = -k_x^2 X(x) \quad (8.12)$$

This equation has exactly the form of the Schrödinger equation (7.51) for a particle in a one-dimensional rigid box,

$$\psi''(x) = -k^2\psi(x) \quad (8.13)$$

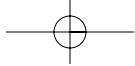
whose solutions we have already discussed in Chapter 7. In particular, we saw that this equation has acceptable solutions only when the constant on the right is negative, which is why we called the constant in (8.11) $-k_x^2$.

An exactly parallel argument shows that the quantity $Y''(y)/Y(y)$ in (8.10) has to be independent of y , that is,

$$\frac{Y''(y)}{Y(y)} = \text{constant} \quad (8.14)$$

or if we call this second constant $-k_y^2$

$$Y''(y) = -k_y^2 Y(y). \quad (8.15)$$



254 Chapter 8 • The Three-Dimensional Schrödinger Equation

We see that the method of separation of variables has let us replace the partial differential equation (8.6), involving the two variables x and y , by two ordinary differential equations (8.12) and (8.15), one of which involves only the variable x , and the other only y .

Before we seek the acceptable solutions of these two equations, we must return to the boundary condition that $\psi(x, y)$ is zero at the edges of our box ($x = 0$ or a , and $y = 0$ or a). Since $\psi(x, y) = X(x)Y(y)$, this requires that

$$X(x) = 0 \quad \text{when } x = 0 \text{ or } a \quad (8.16)$$

and

$$Y(y) = 0 \quad \text{when } y = 0 \text{ or } a. \quad (8.17)$$

The differential equation (8.12) and boundary conditions (8.16) for $X(x)$ are exactly the equation and boundary conditions for a particle in a *one-dimensional* rigid box. And we already know the solutions for that problem: The wave function must have the form

$$X(x) = B \sin k_x x$$

where B is a constant. This satisfies the boundary conditions only if k_x is an integer multiple of π/a ,

$$k_x = \frac{n_x \pi}{a} \quad (8.18)$$

where n_x is any positive integer

$$n_x = 1, 2, 3, \dots$$

Therefore,

$$X(x) = B \sin k_x x = B \sin \frac{n_x \pi x}{a} \quad (8.19)$$

The equation and boundary conditions for $Y(y)$ are also the same as those for a one-dimensional rigid box, and there are acceptable solutions only if

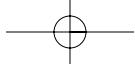
$$k_y = \frac{n_y \pi}{a} \quad (8.20)$$

(with n_y any positive integer), in which case

$$Y(y) = C \sin k_y y = C \sin \frac{n_y \pi y}{a} \quad (8.21)$$

Combining (8.19) and (8.21), we find for the complete wave function

$$\psi(x, y) = X(x)Y(y) = BC \sin k_x x \sin k_y y \quad (8.22)$$

Section 8.3 • The Two-Dimensional Square Box **255**

or

$$\psi(x, y) = A \sin \frac{n_x \pi x}{a} \sin \frac{n_y \pi y}{a} \quad (8.23)$$

where n_x and n_y are any two positive integers.* In writing (8.23), we have renamed the constant BC as A ; the value of this constant is fixed by the normalization condition that the integral of $|\psi|^2$ over the whole box must be 1.

Using the form (8.22) we can see the physical significance of the separation constants k_x and k_y . If we fix y and move in the x direction, then ψ varies sinusoidally in x , with wavelength $\lambda = 2\pi/k_x$. According to de Broglie, this means that the particle has momentum in the x direction of magnitude $h/\lambda = \hbar k_x$. Since a similar argument can be applied in the y direction, we conclude that

$$|p_x| = \hbar k_x \quad \text{and} \quad |p_y| = \hbar k_y \quad (8.24)$$

By analogy with the one-dimensional wave number k , satisfying $p = \hbar k$, we can think of k_x and k_y as the components of a wave vector. Note, however, that since

$$\sin k_x x = \frac{e^{ik_x x} - e^{-ik_x x}}{2i}$$

the wave function (8.22) is a superposition of states with $p_x = \pm \hbar k_x$, and similarly with $p_y = \pm \hbar k_y$. This is the same situation that we encountered in Section 7.4 for the one-dimensional rigid box and explains the absolute value signs in Eq. (8.24).

Allowed Energies

In solving for the wave function (8.22), we have temporarily lost sight of the energy E . In fact, the last place that E appeared was in (8.10):

$$\frac{X''(x)}{X(x)} + \frac{Y''(y)}{Y(y)} = -\frac{2ME}{\hbar^2} \quad (8.25)$$

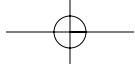
Now, we know from (8.12) that X''/X is the constant $-k_x^2$, and from (8.18) that $k_x = n_x \pi/a$. Therefore, X''/X is equal to $-n_x^2 \pi^2/a^2$. Inserting this and the corresponding expression for Y''/Y , into (8.25), we obtain

$$-\frac{n_x^2 \pi^2}{a^2} - \frac{n_y^2 \pi^2}{a^2} = -\frac{2ME}{\hbar^2}$$

Solving for E , we find that the allowed values of the energy are

$$E = E_{n_x, n_y} = \frac{\hbar^2 \pi^2}{2Ma^2} (n_x^2 + n_y^2) \quad (8.26)$$

*By labeling these two integers n_x and n_y , we do not wish to imply that there is necessarily a vector \mathbf{n} of which n_x and n_y are the components. For the moment, n_x and n_y are simply two integers, one of which characterizes the function $X(x)$ and the other $Y(y)$.



256 Chapter 8 • The Three-Dimensional Schrödinger Equation

where n_x and n_y are any two positive integers. This energy is the sum of two terms, each of which has exactly the form of an allowed energy for the one-dimensional box; namely,

$$E = \frac{\hbar^2\pi^2}{2Ma^2}n^2 \quad [n = 1, 2, 3, \dots] \quad (8.27)$$

If we adopt the notation

$$E_0 = \frac{\hbar^2\pi^2}{2Ma^2} \quad (8.28)$$

where you can think of the subscript 0 as standing for one-dimensional, we can rewrite the allowed energies for a particle in a two-dimensional square box as

$$E = E_{n_x, n_y} = E_0(n_x^2 + n_y^2) \quad (8.29)$$

Quantum Numbers

Just like the one-dimensional box, the two-dimensional box has energy levels that are quantized. The main difference is that where the one-dimensional energy levels are characterized by a single integer n , the two-dimensional levels are given by two integers, n_x , and n_y . We are going to find many more examples of quantities whose allowed values are characterized by integers (and sometimes half integers, such as $\frac{1}{2}, 1\frac{1}{2}, \dots$). In general, any integer or half integer that gives the allowed values of some physical quantity is called a **quantum number**. With this terminology, we can say that the energy levels of a particle in a two-dimensional square box are characterized by two quantum numbers, n_x and n_y .

The lowest possible energy for the two-dimensional box occurs when both quantum numbers are equal to 1,

$$n_x = n_y = 1$$

and the corresponding, ground-state, energy is given by (8.29) as

$$E_{11} = 2E_0$$

The first excited energy occurs when $n_x = 1, n_y = 2$, or vice versa.

$$E_{12} = E_{21} = 5E_0$$

In Fig. 8.2 are sketched the lowest four energy levels for the square box.

FIGURE 8.2

The energy levels of a particle in a two-dimensional, square rigid box. The lowest allowed energy is $2E_0$; the line at $E = 0$ is merely to show the zero of the energy scale. The degeneracies, listed on the right, refer to the number of independent wave functions with the same energy.

n_x	n_y	E_{n_x, n_y}	Degeneracy
1	3	$10E_0$	2
3	1	$10E_0$	2
2	2	$8E_0$	1
1	2	$5E_0$	2
2	1	$5E_0$	2
1	1	$2E_0$	1
		$E = 0$	

Degeneracy

An important new feature of the two-dimensional box is that there can be several different wave functions for which the particle has the same energy. For example, we saw that $E_{12} = E_{21} = 5E_0$. That is, the state with $n_x = 1, n_y = 2$ has the same energy as that with $n_x = 2$ and $n_y = 1$. The corresponding wave functions are

$$\psi_{12} = A \sin \frac{\pi x}{a} \sin \frac{2\pi y}{a} \quad \text{and} \quad \psi_{21} = A \sin \frac{2\pi x}{a} \sin \frac{\pi y}{a} \quad (8.30)$$

Since these correspond to different probability densities $|\psi|^2$, they represent experimentally distinguishable states that happen to have the same energy.

In general, if there are N independent wave functions ($N > 1$), all with the same energy E , we say that the energy level E is **degenerate** and has **degeneracy N** (or is N -fold degenerate). If there is only one wave function with energy E , we say that the energy E is **nondegenerate** (or has degeneracy 1). Looking at Fig. 8.2, we see that the ground state and the second excited state of the square box are nondegenerate, while the first and third excited states are both twofold degenerate. In general, most of the levels $E_{11}, E_{22}, E_{33}, \dots$ are nondegenerate, while most of the levels E_{n_x, n_y} with $n_x \neq n_y$ are twofold degenerate, since $E_{n_x, n_y} = E_{n_y, n_x}$. A few of the levels have higher degeneracies; for example, since

$$5^2 + 5^2 = 1^2 + 7^2$$

it follows that $E_{55} = E_{17} = E_{71}$, and this level is threefold degenerate; since

$$1^2 + 8^2 = 4^2 + 7^2$$

it follows that $E_{18} = E_{81} = E_{47} = E_{74}$, and this level is fourfold degenerate.

In Chapter 10 we will see that degeneracy has an important effect on the structure and chemical properties of atoms. Therefore, it is important not just to find the energy levels of a quantum system, but to find the degeneracy of each level.

Contour Maps of $|\psi|^2$

It is often important to know how the probability density $|\psi(x, y)|^2$ is distributed in space. Because $|\psi(x, y)|^2$ depends on two variables, it is harder to visualize than in the one-dimensional case. One method that is quite successful is to draw a contour map of $|\psi(x, y)|^2$. Figure 8.3 shows such a contour map for the ground-state density

$$|\psi(x, y)|^2 = A^2 \sin^2\left(\frac{\pi x}{a}\right) \sin^2\left(\frac{\pi y}{a}\right) \quad (8.31)$$

The density $|\psi|^2$ is maximum at the center of the box $x = y = a/2$. The contours shown are for $|\psi|^2$ equal to 95%, 50%, and 5% of its maximum value. Notice how the contour lines become more square near the edges of the box. The contour line $|\psi|^2 = 0$ is, of course, the square boundary of the box itself.

Figure 8.4 shows the same three contour lines for each of three excited states. Notice how the higher energies correspond to more rapid oscillations of the wave functions and hence to larger numbers of hills and valleys on the map.

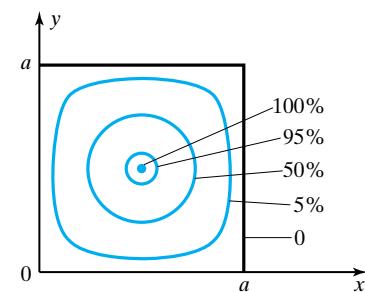
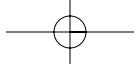


FIGURE 8.3

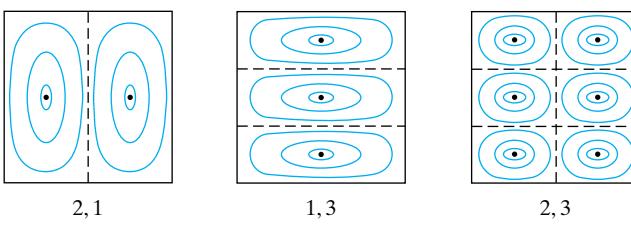
Contour map of the probability density $|\psi|^2$ for the ground state of the square box. The percentages shown give the value of $|\psi|^2$ as a percentage of its maximum value.



258 Chapter 8 • The Three-Dimensional Schrödinger Equation

FIGURE 8.4

Contour maps of $|\psi|^2$ for three excited states of the square box. The two numbers under each picture are n_x and n_y . The dashed lines are nodal lines, where $|\psi|^2$ vanishes; these occur where ψ passes through zero as it oscillates from positive to negative values.



Example 8.2

Having solved the Schrödinger equation for a particle in the two-dimensional square box, one can solve the corresponding three-dimensional problem very easily (see Problem 8.15). The result is that the allowed energies for a mass M in a rigid cubical box of side a have the form

$$E = E_0(n_x^2 + n_y^2 + n_z^2) \quad (8.32)$$

where $E_0 = \hbar^2\pi^2/(2Ma^2)$ is the same energy introduced in (8.28), and the quantum numbers n_x, n_y, n_z are any three positive integers. Use this result to find the lowest five energy levels and their degeneracies for a mass M in a rigid cubical box of side a .

Equation (8.32) shows that the energies of a particle in a three-dimensional box are characterized by three quantum numbers n_x, n_y, n_z . The lowest energy occurs for $n_x = n_y = n_z = 1$ and is

$$E_{111} = 3E_0$$

The next level corresponds to the three quantum numbers being 2, 1, 1 or 1, 2, 1 or 1, 1, 2:

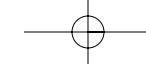
$$E_{211} = E_{121} = E_{112} = 6E_0$$

This level is evidently threefold degenerate. The higher levels are easily calculated, and the first five levels are found to be as shown in Fig. 8.5.

(n_x, n_y, n_z)	E	Degeneracy
(222)	$12E_0$	1
(113) or (131) or (311)	$11E_0$	3
(122) or (212) or (221)	$9E_0$	3
(112) or (121) or (211)	$6E_0$	3
(111)	$3E_0$	1
	$E = 0$	

FIGURE 8.5

The first five levels and their degeneracies for a particle in a three-dimensional cubical box.



8.4 The Two-Dimensional Central-Force Problem

Many physical systems involve a particle that moves under the influence of a **central force**; that is, a force that always points exactly toward, or away from, a force center O . In classical mechanics a famous example of a central force is the force of the sun on a planet. In atomic physics the most obvious example is the hydrogen atom, in which the electron is held to the proton by the central Coulomb force. Other examples where the force is at least approximately central include the motion of any one electron in a multielectron atom, and the motion of either atom as it orbits around the other atom in a diatomic molecule.

If the force on a particle is central, it does no work when the particle moves in any direction perpendicular to the radius vector, as shown in Fig. 8.6. This means that the particle's potential energy U is constant in any such displacement. Thus U may depend on the particle's distance, r , from the force center O , but not on its direction. Therefore, instead of writing the potential energy as $U(x, y, z)$, we can write simply $U(r)$ when the force is central. This property of central forces will allow us to solve the Schrödinger equation using separation of variables.

As an introduction to the three-dimensional central-force problem, we consider first a two-dimensional particle moving in a central-force field. We will not present a complete solution of the Schrödinger equation for this system since it is fairly complicated and we are not really interested in two-dimensional systems here. Nevertheless, we will carry it far enough to see two important facts: First, like the energies for a square box, the allowed energies of a two-dimensional particle in a central-force field are given by two quantum numbers. Second, we will find that one of the two quantum numbers is closely connected with the angular momentum of the particle.

Since the potential energy U depends only on r , the distance of the particle from the force center O , it is natural to adopt r as one of our coordinates. In two dimensions the simplest way to do this is to use **polar coordinates** (r, ϕ) as defined in Fig. 8.7.* It is a simple trigonometric exercise to express x and y in terms of r and ϕ as in Fig. 8.7, or vice versa (Problem 8.16). Note that r is defined as the distance from O to the point of interest and is therefore always positive: $0 \leq r < \infty$. If we increase the angle ϕ by 2π (one complete revolution), we come back to our starting direction. Therefore, you must remember that $\phi = \phi_0$ and $\phi = \phi_0 + 2\pi$ represent exactly the same direction.

The wave function ψ , which depends on x and y , can just as well be expressed as a function of r and ϕ :

$$\psi = \psi(r, \phi)$$

and the Schrödinger equation can similarly be rewritten in terms of r and ϕ . When one rewrites the partial derivatives of the Schrödinger equation in terms of r and ϕ , one finds that

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{\partial^2 \psi}{\partial r^2} + \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \phi^2} \quad (8.33)$$

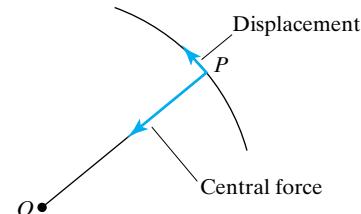


FIGURE 8.6

A central force points exactly toward or away from O . If the particle undergoes a displacement perpendicular to the radius vector OP , the force does no work and the potential energy U is therefore constant.

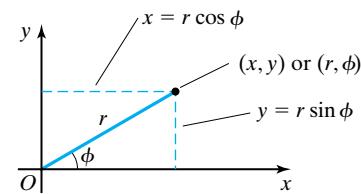
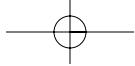


FIGURE 8.7

Definition of the polar coordinates r and ϕ in two dimensions.

*In two dimensions, the angle that we are calling ϕ is more often called θ . However, in three dimensions it is usually called ϕ , and our primary interest will be in three dimensions.



260 Chapter 8 • The Three-Dimensional Schrödinger Equation

If you have had some experience with handling partial derivatives, you should be able to verify this rather messy identity (Problem 8.19). Otherwise, it is probably simplest for now to accept it without proof. In any case, it is helpful to note that all three terms are dimensionally consistent. Given the identity (8.33), we can rewrite the Schrödinger equation (8.6) in terms of r and ϕ as

$$\frac{\partial^2 \psi}{\partial r^2} + \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \phi^2} = \frac{2M}{\hbar^2} [U(r) - E] \psi \quad (8.34)$$

Separation of Variables

The equation (8.34) can be solved by separation of variables, very much as described in Section 8.3, except that we now work with the coordinates r and ϕ instead of x and y . We first seek a solution with the separated form

$$\psi(r, \phi) = R(r)\Phi(\phi) \quad (8.35)$$

Substituting (8.35) into (8.34), we find that

$$\Phi(\phi) \left[R''(r) + \frac{R'(r)}{r} \right] + \frac{R(r)}{r^2} \Phi''(\phi) = \frac{2M}{\hbar^2} [U(r) - E] R(r) \Phi(\phi)$$

where, as before, primes denote differentiation with respect to the argument ($R' = dR/dr$, $\Phi' = d\Phi/d\phi$). If we now multiply both sides by $r^2/(R\Phi)$ and regroup terms, this gives

$$\frac{\Phi''(\phi)}{\Phi(\phi)} = -\frac{r^2 R''(r) + r R'(r)}{R(r)} + \frac{2Mr^2}{\hbar^2} [U(r) - E] \quad (8.36)$$

for all r and ϕ . The equation (8.36) has the form

$$(\text{function of } \phi) = (\text{function of } r), \quad (8.37)$$

and this will allow us to separate variables. Notice that the separation in (8.37) occurs because the potential energy in (8.36) depends only on r (which follows because the force is central). This explains why we went to the trouble of rewriting the equation in terms of r and ϕ . It would not have separated if we had used x and y , since $U(r)$ depends on both x and y (because $r = \sqrt{x^2 + y^2}$).

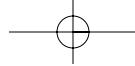
The right side of (8.37) is a function of r but is independent of ϕ . Since the two sides are equal for all r and ϕ , it follows that the left side is *also* independent of ϕ , and hence a *constant*. By a similar argument, the right side is likewise constant, and by (8.37) these two constants are equal. It is traditional (and, as we will see, convenient) to call this constant* $-m^2$. Since each side of (8.36) is equal to the constant $-m^2$, we get two equations:

$$\Phi''(\phi) = -m^2 \Phi(\phi) \quad (8.38)$$

and

$$R'' + \frac{R'}{r} - \left[\frac{m^2}{r^2} + \frac{2M}{\hbar^2} (U - E) \right] R = 0 \quad (8.39)$$

*It is to avoid confusion with this m that we are using M for the mass of the particle.

Section 8.4 • The Two-Dimensional Central-Force Problem **261**

Once again, separation of variables has reduced a single partial differential equation in two variables to two separate equations, each involving just one variable. The two equations (8.38) and (8.39) are called the ϕ equation and the radial equation. We discuss the ϕ equation (8.38) first.

We already know that the general solution of (8.38) is an arbitrary combination of $\cos m\phi$ and $\sin m\phi$, or, equivalently, of $e^{im\phi}$ and $e^{-im\phi}$. We can economize a little in notation if we use the second pair and if we agree to let m be positive or negative. For example, both $e^{i\phi}$ and $e^{-i\phi}$ can be written as $e^{im\phi}$ if we let m be ± 1 . Thus, we can say that all solutions of (8.38) can be built up from the functions

$$\Phi(\phi) = e^{im\phi} \quad (8.40)$$

with m positive or negative.

The ϕ equation (8.38) proved easy to solve, but we must now ask whether there are any boundary conditions to be met. In fact, there are. For any given r , the points labeled by ϕ and $\phi + 2\pi$ are the same. Therefore, the wave function $\psi(r, \phi)$ must satisfy

$$\psi(r, \phi) = \psi(r, \phi + 2\pi)$$

For our separated solutions, this means that

$$\Phi(\phi) = \Phi(\phi + 2\pi)$$

that is, $\Phi(\phi)$ must be periodic and must repeat itself each time ϕ increases by 2π .

It is known from trigonometry that the function $\cos m\phi$ is periodic and repeats itself every 2π if m is an integer, but not otherwise (Fig. 8.8). The same is true of $\sin m\phi$ and hence also of $e^{im\phi} = \cos m\phi + i \sin m\phi$. Thus the solution (8.40) of the ϕ equation is acceptable if and only if m is an integer:

$$m = 0, \pm 1, \pm 2, \dots \quad (8.41)$$

Incidentally, we can now see why it was convenient to use the notation $-m^2$ for the separation constant in (8.38). If we had called it K , for instance, then (8.41) would have read $\sqrt{-K} = 0, \pm 1, \dots$.

Quantization of Angular Momentum

Our conclusion so far is that there are solutions of the Schrödinger equation (8.34) with the form

$$\psi(r, \phi) = R(r)e^{im\phi} \quad (8.42)$$

provided the quantum number m is an integer. That m must be an integer indicates that something related to the ϕ dependence of $\psi(r, \phi)$ is quantized. To decide what this is, let us fix r so that we can study just the ϕ dependence. If we let ϕ increase, we move around a circle, as shown in Fig. 8.9, the distance that we travel being $s = r\phi$. According to (8.42), the wave function varies sinusoidally with the distance s :

$$\psi(r, \phi) \propto e^{im\phi} = e^{i(m/r)s} \quad (8.43)$$

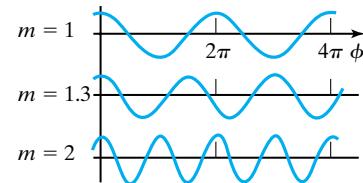


FIGURE 8.8

If m is an integer, the function $\cos m\phi$ repeats itself each time ϕ increases by 2π ; for intermediate values, it does not.

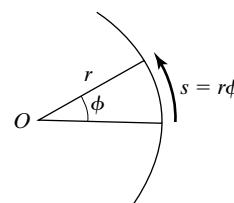
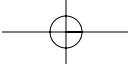


FIGURE 8.9

If we move through an angle ϕ on a circle of radius r , the distance traveled is $s = r\phi$.



262 Chapter 8 • The Three-Dimensional Schrödinger Equation

We saw in Chapter 7 that a wave function

$$e^{ikx} \quad (8.44)$$

where x is the distance along a line, represents a particle with momentum $\hbar k$ along that line.* Comparison of (8.43) and (8.44) suggests that (8.43) represents a particle with momentum

$$p_{\text{tang}} = \hbar \frac{m}{r}$$

in the direction tangential to the circle. If we multiply p_{tang} by r , this gives the angular momentum

$$L = p_{\text{tang}} r = m\hbar$$

that is, the wave functions (8.42) define states in which the particle has a definite angular momentum

$$L = m\hbar \quad (8.45)$$

That m has to be an integer shows that L is quantized in multiples of \hbar , as originally proposed by Bohr. That is, we have justified Bohr's quantization of angular momentum.

You should bear in mind that we have so far discussed the central-force problem only in two dimensions. For a particle confined to the x - y plane, the angular momentum L is the same thing as its z component L_z . Thus, it is not clear whether the result (8.45) will apply to L or L_z when we go on to discuss three-dimensional motion. In fact, we will see that in three dimensions it is L_z that is restricted to integer multiples of \hbar :

$$L_z = m\hbar \quad [m = 0, \pm 1, \pm 2, \dots] \quad (8.46)$$

Of course, there is nothing special about the z axis (in three dimensions), and the general statement of (8.46) is that *any* component of the vector \mathbf{L} is restricted to integer multiples of \hbar .

The Energy Levels

Let us turn now to the radial equation (8.39),

$$R'' + \frac{R'}{r} - \left[\frac{m^2}{r^2} + \frac{2M}{\hbar^2} (U - E) \right] R = 0 \quad (8.47)$$

Since this equation contains the energy E , its solution will determine the allowed values of E . The details of the solution depend on the particular potential-energy function $U(r)$, which we have not specified. However, we can understand several general features: The radial equation is an ordinary

*Recall that the full time-dependent wave function corresponding to (8.44) is $\Psi(x, t) = \exp[i(kx - \omega t)]$. This wave travels to the right and is sinusoidal in x with wavelength $2\pi/k$, that is, it has momentum $p = h/\lambda = \hbar k$ to the right.

differential equation, which involves the energy E as a parameter. Just as with the one-dimensional Schrödinger equation, one can show that there are acceptable solutions only for certain particular values of E , and these are the allowed energies of our particle. Notice that the equation (8.47) that determines the allowed energies depends on the quantum number m . Thus for each value of m , we have a different equation to solve and will usually get different allowed energies; that is, the allowed energies of our particle will depend on its angular momentum. This is what one would expect classically: The more angular momentum the particle has, the more kinetic energy it will have in its orbital motion; thus, in general, we expect different energies for different angular momenta.

For each value of m , we could imagine finding all the allowed energies. We could then label them in increasing order by an integer $n = 1, 2, 3, \dots$, so that the n th level with angular momentum $m\hbar$ would have energy $E_{n,m}$, as shown in Fig. 8.10. Just as with the square box, each level is then identified by two quantum numbers. In this case one of the quantum numbers, m , identifies the angular momentum, while the other, n , identifies the energy level for given m .

The radial equation (8.47) involves m only in the term m^2/r^2 . Because this depends on m^2 , we get the same equation, and hence the same energies, whether m is positive or negative.

$$E_{n,m} = E_{n,-m}$$

In other words, except when $m = 0$ there are two states with the same energy, and the level $E_{n,m}$ is twofold degenerate. This is a property that we would also have found in a classical analysis: Two states that differ only by having $L_z = \pm m\hbar$ are different only because the particle is orbiting in opposite directions, and we would expect two such states to have the same energy.

We have not actually *found* the allowed energies $E_{n,m}$ of our two-dimensional particle. These depend on the particular potential-energy function $U(r)$ under consideration. Whatever the form of $U(r)$, the detailed solution of the radial equation (8.47) is fairly complicated. Since our real concern is with three-dimensional systems, we will not pursue the two-dimensional problem any further here.

8.5 The Three-Dimensional Central-Force Problem

The three-dimensional central-force problem — the problem of finding the motion of a particle subject to a central force in three dimensions — is perhaps the single most important problem in all of quantum mechanics. The motion of an electron in a hydrogen atom — in some ways the most important of atoms — is an example. Our understanding of all atoms, and ultimately of all chemistry, depends on an understanding of the central-force problem; and the same can be said of the internal motion of atomic nuclei, the main theme of nuclear physics.

Obviously, it is crucial that you get a good grasp of the three-dimensional central-force problem. In principle, at least, the three-dimensional problem is similar to the two-dimensional. Unfortunately, the former involves some complicated, and frankly quite ugly, mathematics. Thus, the next few sections are going to be quite heavy going and will contain more statements that “it can be shown” than we would like. We can only assure you that this hard work is in a

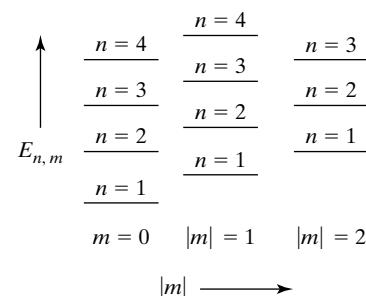
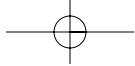


FIGURE 8.10

General appearance of the energy levels for a two-dimensional central force. Energy is plotted upward and the angular-momentum quantum number is plotted to the right. For each value of m , there may be several possible energies, which we label with a quantum number $n = 1, 2, 3, \dots$. For each pair of values n and m , we denote the corresponding energy by $E_{n,m}$. Notice that the variable plotted horizontally is the *absolute value* of m , since $E_{n,m}$ depends only on the magnitude of the angular momentum, not its direction.



264 Chapter 8 • The Three-Dimensional Schrödinger Equation

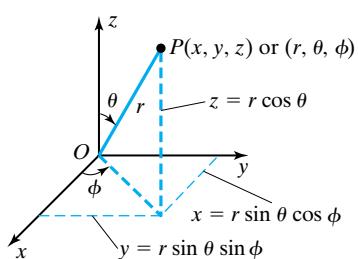


FIGURE 8.11

The spherical polar coordinates of a point P are (r, θ, ϕ) , where r is the distance OP , θ is the angle between OP and the z axis, and ϕ is the angle between the xz plane and the vertical plane containing OP .

good cause, namely, the understanding of one of the historic achievements of twentieth-century science.

Since the potential energy of any central force depends only on r (the particle's distance from the force center O), we first choose a coordinate system that includes r as one of the coordinates. For this, we use **spherical polar coordinates**, which are defined in Fig. 8.11. Any point P is identified by the three coordinates (r, θ, ϕ) , where r is the distance from O to P , θ is the angle between the z -axis and OP , and ϕ is the angle between the xz plane and the vertical plane containing OP , as shown.* If we imagine P to be a point on the earth's surface and put the origin O at the earth's center, with the z -axis pointing to the north pole, then θ is the colatitude of P (the latitude measured down from the north pole) and ϕ is its longitude measured in an easterly direction from the xz plane. The angle θ lies between 0, at the north pole, and π at the south pole. If ϕ increases from 0 to 2π (with r and θ fixed), then P circles the earth at fixed latitude and returns to its starting point. The rectangular coordinates (x, y, z) are given in terms of (r, θ, ϕ) by

$$x = r \sin \theta \cos \phi \quad y = r \sin \theta \sin \phi \quad z = r \cos \theta$$

You should verify these expressions for yourself and derive the corresponding expressions for (r, θ, ϕ) in terms of (x, y, z) (Problem 8.20).

To write down the Schrödinger equation in terms of spherical coordinates, we must write the derivatives with respect to x, y, z in terms of r, θ , and ϕ . When this is done, one finds that

$$\begin{aligned} \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} &= \frac{1}{r} \frac{\partial^2}{\partial r^2} (r\psi) \\ &+ \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \end{aligned} \quad (8.48)$$

The proof of this identity is analogous to that of the two-dimensional identity (8.33) (see Problem 8.19); however, it is appreciably more complicated and is certainly not worth giving here. We will simply accept the identity (8.48) and use it to write down the three-dimensional Schrödinger equation in spherical coordinates.

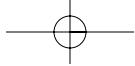
$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (r\psi) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} = \frac{2M}{\hbar^2} [U(r) - E]\psi \quad (8.49)$$

Separation of Variables

The three-dimensional Schrödinger equation for a central force (8.49) can be solved by separation of variables. We start by seeking a solution with the separated form

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi) \quad (8.50)$$

* You should be aware of the horrible historical accident that most mathematicians use θ and ϕ for what most physicists call ϕ and θ .

Section 8.6 • Quantization of Angular Momentum **265**

If we substitute (8.50) into (8.49), we can rearrange the resulting equation in the form

$$\frac{\Phi''(\phi)}{\Phi(\phi)} = (\text{function of } r \text{ and } \theta) \quad (8.51)$$

(For guidance in checking this and the next few steps, see Problem 8.22.) By the now-familiar argument, each side of this equation is equal to the same constant, which we call $-m^2$. This gives us two equations:

$$\Phi''(\phi) = -m^2\Phi(\phi) \quad (8.52)$$

and a second equation involving r and θ . This second equation can next be rearranged in the form

$$(\text{function of } \theta) = (\text{function of } r)$$

Once again each side must be equal to a constant, which we will temporarily call $-k$. This gives two final equations with the form (Problem 8.22)

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \left(k - \frac{m^2}{\sin^2 \theta} \right) \Theta = 0 \quad (8.53)$$

and

$$\frac{d^2}{dr^2}(rR) = \frac{2M}{\hbar^2} \left[U(r) + \frac{k\hbar^2}{2Mr^2} - E \right] (rR) \quad (8.54)$$

We see that separation of variables has reduced the partial differential equation (8.49) in r , θ , and ϕ to three ordinary differential equations, each involving just one variable. Notice that the ϕ equation (8.52) is exactly the same as Equation (8.38) for the two-dimensional case. Notice also that neither the ϕ equation (8.52) nor the θ equation (8.53) involves the potential-energy function $U(r)$. This means that the solutions for the angular functions $\Theta(\theta)$ and $\Phi(\phi)$ will apply to any central-force problem. Finally, note that both $U(r)$ and E appear only in the radial equation (8.54). Therefore, it is the radial equation that determines the allowed values of the energy, and these do depend on the potential-energy function $U(r)$, as one would expect.

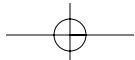
8.6 Quantization of Angular Momentum

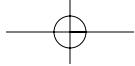
In this section we discuss the two angular equations (8.52) and (8.53) that resulted from separating the three-dimensional Schrödinger equation. The first of these is exactly the ϕ equation that arose in the two-dimensional central-force problem, and it has the same solutions

$$\Phi(\phi) = e^{im\phi} \quad (8.55)$$

Since $\Phi(\phi)$ must be periodic with period 2π , it follows, as before, that m must be an integer

$$m = 0, \pm 1, \pm 2, \dots$$





266 Chapter 8 • The Three-Dimensional Schrödinger Equation

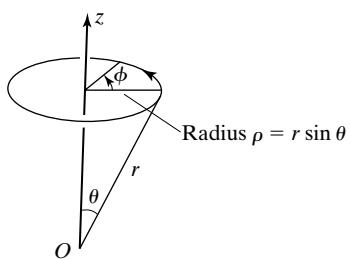


FIGURE 8.12

If we fix r and θ and let ϕ vary, we move around a circle of radius $\rho = r \sin \theta$.

The significance of m is essentially the same as in two dimensions: If we fix r and θ and let ϕ vary, then we move around a circle about the z -axis. The radius of this circle is $\rho = r \sin \theta$, as shown in Fig. 8.12. In terms of the distance s traveled around this circle, the angle ϕ is

$$\phi = \frac{s}{\rho}$$

and we can temporarily rewrite (8.55) as

$$\Phi(\phi) = e^{im\phi} = e^{i(m/\rho)s} \quad (8.56)$$

Comparing this with the familiar one-dimensional wave e^{ikx} with momentum $\hbar k$, we see that (8.56) represents a state with tangential momentum

$$p_{\text{tang}} = \frac{\hbar m}{\rho} \quad (8.57)$$

If we multiply p_{tang} by the radius ρ , we obtain the z component of angular momentum, $L_z = p_{\text{tang}}\rho$. Thus, from (8.57) our wave function represents a particle with

$$L_z = m\hbar \quad [m = 0, \pm 1, \pm 2, \dots]$$

as anticipated in Section 8.4.

The θ Equation

The second angular equation (8.53), which determines $\Theta(\theta)$, is much harder and we will have to be satisfied with stating its solutions. The equation is one of the standard equations of mathematical physics and is called Legendre's equation. It has solutions for any value of the separation constant k . However, for most values of k , these solutions are infinite at $\theta = 0$ or at $\theta = \pi$ and are therefore physically unacceptable. It turns out that the equation has one (and only one) acceptable solution for each k of the form

$$k = l(l + 1) \quad (8.58)$$

where l is a positive integer greater than or equal in magnitude to m ,

$$l \geq |m| \quad (8.59)$$

If we denote these acceptable solutions by $\Theta_{lm}(\theta)$, our solutions of the Schrödinger equation have the form

$$\psi(r, \theta, \phi) = R(r)\Theta_{lm}(\theta)e^{im\phi} \quad (8.60)$$

We will find the specific form of the function $\Theta_{lm}(\theta)$ for a few values of l and m later. Figure 8.13 shows the function $\Theta_{lm}(\theta)$ for the case $l = 2$ and $m = 0$, as well as one of the unacceptable solutions for the case $l = 1.75$, $m = 0$.

The physical significance of the quantum number m is, as we have seen, that a particle with the wave function (8.60) has a definite value of L_z equal to $m\hbar$.

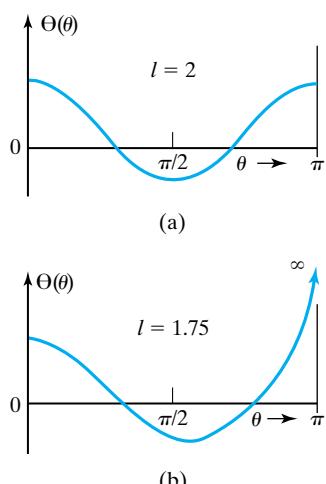
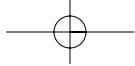


FIGURE 8.13

- (a) If the constant k has the form $l(l + 1)$ with l an integer greater than or equal to $|m|$, the θ equation (8.53) has one acceptable solution, finite for all θ from 0 to π . The picture shows this acceptable solution for the case $m = 0$, $l = 2$.
- (b) Otherwise, every solution of the θ equation is infinite at $\theta = 0$ or π or both. The picture shows a solution that is finite at $\theta = 0$ but infinite at $\theta = \pi$ for the case $m = 0$ and $l = 1.75$.

Section 8.6 • Quantization of Angular Momentum **267**

In more advanced texts it is shown that a particle with the wave function (8.60) also has a definite value for the magnitude of \mathbf{L} equal to

$$L = \sqrt{l(l+1)\hbar} \quad (8.61)$$

That is, the quantum number l identifies the magnitude of \mathbf{L} , according to (8.61).

The quantum number m can be any integer (positive or negative), while for given m , l can be any integer greater than or equal to the magnitude of m . Turning this around, we can say that l can be any positive integer

$$l = 0, 1, 2, \dots \quad (8.62)$$

while for given l , m can be any integer less than or equal (in magnitude) to l . That is,

$$L_z = m\hbar, \quad m = l, l-1, \dots, -l \quad (8.63)$$

According to (8.61) and (8.62) the possible magnitudes of \mathbf{L} are as follows:

Quantum number, l :	0	1	2	3	4	...
Magnitude:	0	$\sqrt{2}\hbar$	$\sqrt{6}\hbar$	$\sqrt{12}\hbar$	$\sqrt{20}\hbar$...

When l is large we can approximate $l(l+1)$ by l^2 and write

$$L \approx l\hbar$$

Thus, for large l the possible magnitudes of \mathbf{L} are close to those of the Bohr model, but for small l , there is an appreciable difference. (See Problems 8.27 and 8.28.)

The Vector Model

We have found wave functions for which the magnitude and z component of \mathbf{L} are quantized. This is, of course, a purely quantum result. Nevertheless, it is sometimes useful to try to visualize it classically. The magnitude of \mathbf{L} is $\sqrt{l(l+1)\hbar}$, so we imagine a vector of length $L = \sqrt{l(l+1)\hbar}$. Since $L_z = m\hbar$, this vector must be oriented so that its z component is $m\hbar$, and since m can take any of the $(2l+1)$ different values (8.63), there are $(2l+1)$ possible orientations, as shown in Fig. 8.14 for the case $l = 2$. We can describe this state of affairs by saying that the spatial orientation of \mathbf{L} is quantized, and in the older literature the quantization of L_z was sometimes called “space quantization.”

We should emphasize that there is nothing special about the z axis. When we defined our spherical coordinates we chose to use the z direction as the polar axis ($\theta = 0$); and when we separated variables, the Schrödinger equation led us to wave functions with a definite value for L_z . If we had chosen the x direction as the polar axis, the same procedure would have produced states with a definite value for L_x , and so on. For the moment it makes no difference which component of \mathbf{L} we choose to focus on, and we will continue to work with states that have definite L_z . However, in more advanced books it is shown that the Heisenberg uncertainty principle extends to angular momentum and implies that no two components of \mathbf{L} can simultaneously have definite values

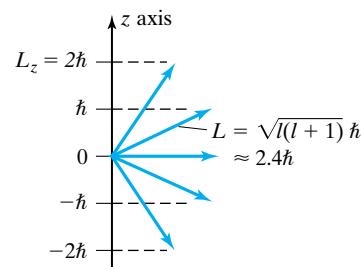
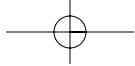


FIGURE 8.14

Classical representation of the quantized values of angular momentum \mathbf{L} for the case $l = 2$. The z component has $(2l+1) = 5$ possible values, $L_z = m\hbar$ with $m = 2, 1, 0, -1, -2$. The magnitude of \mathbf{L} is $L = \sqrt{l(l+1)\hbar} = \sqrt{2 \times 3}\hbar \approx 2.4\hbar$ in all five cases.



268 Chapter 8 • The Three-Dimensional Schrödinger Equation

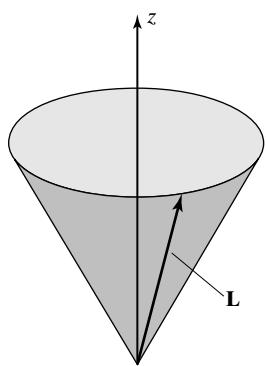


FIGURE 8.15

The quantum properties of angular momentum can be visualized by imagining the vector \mathbf{L} randomly distributed on the cone shown. This represents the quantum situation, where L and L_z have definite values but L_x and L_y do not.

(except in the special case that all three components are zero). Therefore, states that have definite L_z , do not have definite values of L_x or L_y .*

Since our wave functions do not have definite values of L_x and L_y , the vectors shown in Fig. 8.14, with definite components in all directions, are a bit misleading. One must somehow imagine that the components L_x and L_y are random and no longer have definite values. Since the magnitude and z component are fixed, this means that the vector \mathbf{L} is randomly distributed on a cone as shown in Fig. 8.15. This reflects the quantum situation, where L_x and L_y simply do not have definite values. It is sometimes helpful to use this classical picture — often called the **vector model** — as an aid in thinking about the quantum properties of angular momentum.

The wave functions that we have found for a particle in a central-force field have the form

$$\psi(r, \theta, \phi) = R(r)\Theta_{lm}(\theta)e^{im\phi} \quad (8.64)$$

and represent a particle with definite values for the magnitude and z component of \mathbf{L} :

$$L = \sqrt{l(l+1)\hbar} \quad \text{and} \quad L_z = m\hbar$$

It is sometimes important to know the explicit form of these wave functions. Fortunately, we will usually be concerned only with states with small values of l — in chemistry, for example, all of the electrons most involved in molecular bonding have $l = 0$ or 1 — and for these, the angular wave functions are quite simple, as the following example shows.

Example 8.3

Write down the θ equation (8.53) for the cases that $l = 0$ and that $l = 1$, $m = 0$. Find the angular functions $\Theta_{lm}(\theta)e^{im\phi}$ explicitly for these two cases.

The θ equation (8.53) is

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \left(l(l+1) - \frac{m^2}{\sin^2 \theta} \right) \Theta = 0 \quad (8.65)$$

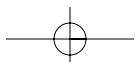
[Recall that the separation constant k got renamed $l(l+1)$.] If $l = 0$, the only allowed value for m is $m = 0$, and the θ equation reduces to

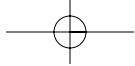
$$\frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) = 0$$

By inspection, we see that one solution of this equation is

$$\Theta(\theta) = \text{constant} \quad (8.66)$$

*The wave functions that have definite L_x are different from those with definite L_z . However, any one of the former can be expressed as a sum of the latter and vice versa. In this sense, it is enough to consider just those with definite L_z . We will see one example of this in Section 8.8.



Section 8.6 • Quantization of Angular Momentum **269**

As stated in connection with (8.58), the θ equation has only one acceptable solution for each value of l and m . (This is illustrated in Problem 8.29.) Therefore, we need look no further for any other solutions. We see that with $l = 0$ the function $\Theta(\theta)$ is independent of θ . Further, with $m = 0$, $e^{im\phi} = 1$ is independent of ϕ . Thus with $l = m = 0$, the wave function (8.64) is actually independent of θ and ϕ and depends only on r .

$$\psi(r, \theta, \phi) = R(r) \times \text{constant} \quad [\text{for } l = m = 0]$$

This means that the probability distribution $|\psi(r, \theta, \phi)|^2$ for a particle with zero angular momentum is spherically symmetric. We will find, for example, that the ground state of the electron in the hydrogen atom has $l = 0$ and hence that the hydrogen atom is spherically symmetric in its ground state.

If $l = 1$, then m can be $m = 1, 0$, or -1 . In this example we are asked to consider the case $m = 0$, for which the θ equation (8.65) reads

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + 2\Theta = 0$$

By inspection, we see that the solution of this equation is

$$\Theta(\theta) = \cos \theta \quad (8.67)$$

Therefore, with $l = 1$ and $m = 0$, the complete wave function given by (8.64) is

$$\psi(r, \theta, \phi) = R(r) \cos \theta \quad [l = 1, m = 0] \quad (8.68)$$

As is obviously the case whenever $m = 0$, this wave function is independent of the angle ϕ . On the other hand, it does depend on θ . Since $|\psi|^2 \propto \cos^2 \theta$, a particle with $l = 1$ and $m = 0$ is most likely to be found near the polar axes $\theta = 0$ and π (where $\cos^2 \theta = 1$), and has zero probability of being found in the x - y plane (where $\theta = \pi/2$ and $\cos \theta = 0$). We will find that this distribution of electrons in atoms has important implications for the shape of many molecules.

The complete angular functions $\Theta_{lm}(\theta)e^{im\phi}$ are called **spherical harmonics** and are usually denoted by

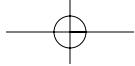
$$Y_{lm}(\theta, \phi) = \Theta_{lm}(\theta)e^{im\phi} \quad (8.69)$$

The functions $\Theta_{lm}(\theta)$ are called **associated Legendre functions** (within a constant multiplicative factor). Table 8.1 lists the first few of these. Notice that we

TABLE 8.1

The first few angular functions $\Theta_{l,m}(\theta)$. The functions with m negative are given by $\Theta_{l,-m} = (-1)^m \Theta_{l,m}$.

$l = 0$	$l = 1$	$l = 2$
$m = 0$	$\sqrt{1/4\pi}$	$\sqrt{3/4\pi} \cos \theta$
$m = 1$	$-\sqrt{3/8\pi} \sin \theta$	$-\sqrt{15/8\pi} \sin \theta \cos \theta$
$m = 2$		$\sqrt{15/32\pi} \sin^2 \theta$



270 Chapter 8 • The Three-Dimensional Schrödinger Equation

have included only the functions with $m \geq 0$ since the θ equation is the same for any m and $-m$, so that $\Theta_{l,-m}(\theta)$ is proportional to $\Theta_{l,m}(\theta)$. [This being the case, it would seem natural to define the functions so that $\Theta_{l,-m} = \Theta_{l,m}$, but the most popular convention has $\Theta_{l,-m} = (-1)^m \Theta_{l,m}$.] The ugly square-root factors are normalization factors. (See Problems 8.33 and 8.34.)

8.7 The Energy Levels of the Hydrogen Atom

Of the three equations that resulted from separating the Schrödinger equation, we have now discussed the two angular ones. It remains to consider the radial equation (8.54), which we rewrite as

$$\frac{d^2}{dr^2}(rR) = \frac{2M}{\hbar^2} \left[U(r) + \frac{l(l+1)\hbar^2}{2Mr^2} - E \right](rR) \quad (8.70)$$

[Remember that the separation constant k in (8.54) got renamed $l(l+1)$.] This is the equation that determines the allowed values of the energy E , which will, of course, depend on the precise form of the potential-energy function $U(r)$. Since the equation involves the angular-momentum quantum number l , the allowed values of E will generally depend on l as well. Notice, however, that (8.70) does not involve the quantum number m . Thus the allowed values of E will not depend on m ; that is, for a given magnitude of \mathbf{L} equal to $\sqrt{l(l+1)}\hbar$, we will find the same allowed energies for all $(2l+1)$ different orientations given by $m = l, l-1, \dots, -l$. This is just what we would expect classically: Since the force field is *spherically symmetric*, the energy of the particle cannot depend on the *orientation* of its orbit. Quantum mechanically it means that in any central-force problem, a level with $L = \sqrt{l(l+1)}\hbar$ will always be at least $(2l+1)$ -fold degenerate. As we will see shortly, it can sometimes be more degenerate since two states with different l may happen to have the same energy.

The detailed solution of the differential equation (8.70) depends on the potential-energy function $U(r)$. As a first and very important example, we consider the electron bound to a proton in a hydrogen atom, for which

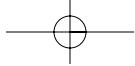
$$U(r) = \frac{-ke^2}{r} \quad (8.71)$$

where k is the Coulomb force constant, $k = 1/4\pi\epsilon_0$. If we substitute (8.71) into (8.70), we obtain the differential equation

$$\frac{d^2}{dr^2}(rR) = \frac{2m_e}{\hbar^2} \left[-\frac{ke^2}{r} + \frac{l(l+1)\hbar^2}{2m_e r^2} - E \right](rR) \quad (8.72)$$

where we have replaced M by m_e , the mass of the electron. This equation has been studied extensively by mathematical physicists. Here we must be content with simply stating the facts about its solutions (but see Problems 8.39 and 8.40 for some simple special cases): The equation (8.72) has acceptable solutions only if E has the form

$$E = -\frac{m_e(k e^2)^2}{2\hbar^2} \frac{1}{n^2} \quad (8.73)$$

Section 8.7 • The Energy Levels of the Hydrogen Atom **271**

where n is any integer greater than l . That is, the energy is quantized, and its allowed values are given by (8.73). You may recognize the first factor in (8.73) as the Rydberg energy, originally defined in (5.22),

$$E_R = \frac{m_e(k e^2)^2}{2\hbar^2} = 13.6 \text{ eV}$$

Thus, solution of the three-dimensional Schrödinger equation for a hydrogen atom has brought us back to exactly the energy levels

$$E = -\frac{E_R}{n^2} \quad (8.74)$$

predicted by the Bohr model. Since these levels are known to be correct, this is a most satisfactory result.

The possible values of the quantum number l are the integers $l = 0, 1, 2, \dots$, and for each value of l , we have stated, the radial equation has a solution only if n is an integer greater than l

$$n > l$$

Turning these statements around, we can say that the possible values of n are the positive integers,

$$n = 1, 2, 3, \dots \quad (8.75)$$

and that, for each value of n , l can be any integer less than n ,

$$l < n$$

that is,

$$l = 0, 1, 2, \dots, (n - 1) \quad (8.76)$$

For the ground state, $n = 1$, the only possible value of l is $l = 0$, and the ground state of hydrogen therefore has zero angular momentum. With $l = 0$, the only possible value of m is $m = 0$, and the ground state is characterized by the unique set of quantum numbers

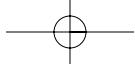
$$\text{ground state: } n = 1, \quad l = 0, \quad m = 0$$

Notice that although the Schrödinger equation and the Bohr model give the same *energy* for the ground state, there is an important difference: Whereas the Bohr model assumed a magnitude $L = 1\hbar$ in the ground state, the Schrödinger equation predicts that $L = 0$, a prediction that is borne out by experiment.

For the first excited level, $n = 2$, there are two possible values of l , namely 0 or 1. If $l = 0$, then m can only be 0; but with $l = 1$, there are three possible orientations of \mathbf{L} , given by $m = 1, 0$, or -1 . Thus, there are four independent wave functions for the first excited level, with quantum numbers:

$$\text{first excited level: } n = 2, \quad l = \begin{cases} 0 & m = 0 \\ \text{or} \\ 1 & m = 1, 0, \text{ or } -1 \end{cases}$$

This means that the first excited level is *fourfold degenerate*.



272 Chapter 8 • The Three-Dimensional Schrödinger Equation

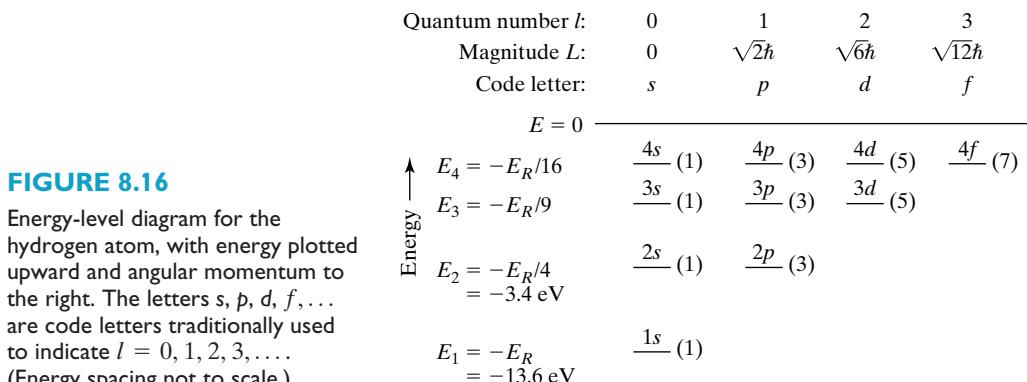


FIGURE 8.16

Energy-level diagram for the hydrogen atom, with energy plotted upward and angular momentum to the right. The letters s, p, d, f, \dots are code letters traditionally used to indicate $l = 0, 1, 2, 3, \dots$ (Energy spacing not to scale.)

For the n th level, there are n possible values of L , given by $l = 0, 1, \dots, (n - 1)$. To display this graphically, it is convenient to draw energy-level diagrams in which the energy is plotted upward as usual, but with the different angular momenta $L = \sqrt{l(l + 1)\hbar}$ shown separately by plotting l horizontally. The first four levels of the hydrogen atom are plotted in this way in Fig. 8.16.

In Figure 8.16 we have introduced the code letters s, p, d, f, \dots , which are traditionally used to identify the magnitude of the angular momentum. These are as follows:

code letter:	s	p	d	f	g	h	i
quantum number l :	0	1	2	3	4	5	6

These code letters are a survival from early attempts to classify spectral lines; in particular, s, p, d , and f stood for *sharp*, *principal*, *diffuse*, and *fundamental*. After f , the letters continue alphabetically, although code letters are seldom used for values of l greater than 6. When specifying the values of n and l , it is traditional to give the number n followed by the code letter for l . Thus the ground state of hydrogen is called $1s$; the first excited level can be $2s$ or $2p$, and so on. Lower case letters, s, p, d, \dots , are generally used when discussing a single electron, and capitals, S, P, D, \dots , when discussing the total angular momentum of a multielectron atom.

Even when n and l are specified, there are still $(2l + 1)$ distinct states corresponding to the $(2l + 1)$ orientations $m = l, l - 1, \dots, -l$. For s states ($l = 0$), there is just one orientation; for p states ($l = 1$) there are $(2 \times 1) + 1 = 3$; for d states ($l = 2$) there are $(2 \times 2) + 1 = 5$, and so on. These numbers are shown in parentheses on the right of each horizontal bar in Fig. 8.16. The total degeneracy of any level can be found by adding all of these numbers for the level in question. For example, the $n = 1$ level is nondegenerate; the $n = 2$ level has degeneracy 4; the $n = 3$ level 9. The n th level has $l = 0, 1, \dots, (n - 1)$, and hence has degeneracy* (Problem 8.35)

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2 \quad (8.77)$$

*Actually, the total degeneracy is twice this answer. This is because the electron has another degree of freedom, called spin, which can be thought of as the angular momentum due to its spinning on its own axis (much as the earth spins on its north-south axis). This spin can have two possible orientations, and for each of the states described here, there are really two states, one for each orientation of the spin. This will be discussed in Chapter 9.

In summary, the stationary states of hydrogen can be identified by three quantum numbers, n , l , and m . The numbers l and m characterize the magnitude and z component of the angular momentum \mathbf{L} . The number n determines the energy as $E_n = -E_R n^2$ and, for this reason, is often called the **principal quantum number**. It is a peculiarity of the hydrogen atom that the energy depends only on n and is independent of l . We will see that in other atoms the energy of an electron is determined mainly by n , but does, nonetheless, depend on l as well.

8.8 Hydrogenic Wave Functions

In many applications of atomic physics it is important to know at least the qualitative behavior of the electron wave functions. In this section we discuss the wave functions for the lowest few levels in the hydrogen atom.

The Ground State

The ground state is the $1s$ state with $n = 1$ and $l = 0$. Since $l = 0$, m has to be zero and, as discussed below (8.66), the wave function is spherically symmetric (that is, it is independent of θ and ϕ and depends only on r).

$$\psi_{1s}(r, \theta, \phi) = R_{1s}(r) \quad (8.78)$$

The radial function $R_{1s}(r)$ is determined by the radial equation (8.72), which we can rewrite (for the particular case that $l = 0$) as

$$\frac{d^2}{dr^2}(rR) = \frac{2m_e}{\hbar^2} \left[-\frac{ke^2}{r} + \frac{E_R}{n^2} \right] (rR) \quad (8.79)$$

If we recall that $\hbar^2/(m_e ke^2)$ is the Bohr radius a_B and that $E_R = ke^2/2a_B$, we can rewrite this equation more simply as

$$\frac{d^2}{dr^2}(rR) = \left(\frac{1}{n^2 a_B^2} - \frac{2}{a_B r} \right) (rR) \quad (8.80)$$

For the case $n = 1$, it is easy to verify that the solution of this equation is (Problem 8.39)

$$R_{1s}(r) = A e^{-r/a_B} \quad (8.81)$$

This wave function is plotted in Fig. 8.17. Since $|\psi|^2$ is the probability density for the electron, it is clear from this picture that the probability density is maximum at the origin. In fact, it is characteristic of all s states (states with zero angular momentum) that $|\psi|^2$ is nonzero at the origin; whereas for any state with $l \neq 0$, $|\psi|^2$ is zero at the origin. This situation is easy to understand classically: A classical particle can be found at $r = 0$ only if its angular momentum is zero. This difference between states with $l = 0$ and those with $l > 0$ has important consequences in multielectron atoms, as we discuss in Chapter 10. It also means that the exact energy of s states is slightly dependent on the spatial extent of the nucleus. In fact, careful measurements of energies of atomic electrons in s states have been used to measure nuclear radii.

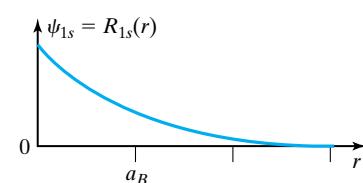
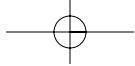


FIGURE 8.17

The wave function (8.81) for the ground state of hydrogen, as a function of r .



274 Chapter 8 • The Three-Dimensional Schrödinger Equation

Since the electron's potential energy depends only on its distance from the nucleus, it is often more important to know the probability of finding the electron at any particular *distance* from the nucleus than to know the probability of its being at any specific position. More precisely, we seek the probability of finding it anywhere between the distances r and $r + dr$ from 0, that is, anywhere in a spherical shell between the radii r and $r + dr$. This can be evaluated if we recall that the probability of finding the electron in a small volume dV is $|\psi|^2 dV$. The volume of this spherical shell is the area of the sphere, $4\pi r^2$, times its thickness, dr .

$$(\text{volume between } r \text{ and } r + dr) = 4\pi r^2 dr \quad (8.82)$$

For the ground state of hydrogen, the wave function depends on r only and is the same at all points in this thin shell. Therefore, the required probability is just

$$P(\text{between } r \text{ and } r + dr) = |\psi|^2 dV = |R(r)|^2 4\pi r^2 dr$$

We can rewrite this as

$$P(\text{between } r \text{ and } r + dr) = P(r) dr \quad (8.83)$$

if we introduce the **radial probability density** (or radial distribution)

$$P(r) = 4\pi r^2 |R(r)|^2 \quad (8.84)$$

We have dropped the subscripts 1s in these important relations since they are in fact true for all wave functions.

An important feature of the function (8.84) is the factor of r^2 , which comes from the factor $4\pi r^2$ in the volume of the spherical shell (8.82). It means that when we discuss the probability of different distances r (as opposed to different positions), large distances are more heavily weighted, just because larger r corresponds to larger spherical shells, with more volume than those with small r .

For the ground state of hydrogen, with wave function (8.81), the radial probability density is

$$P_{1s}(r) = 4\pi A^2 r^2 e^{-2r/a_B} \quad (8.85)$$

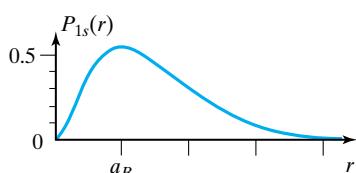


FIGURE 8.18

The probability of finding the electron a distance r from the nucleus is given by the radial probability density $P(r)$. For the 1s or ground state of hydrogen $P(r)$ is maximum at $r = a_B$. The density $P(r)$ has the dimensions of inverse length and is shown here in units of $1/a_B$.

This is plotted in Fig. 8.18.* Perhaps its most striking property is that its maximum is at $r = a_B$. That is, the most probable distance between the electron and proton in the 1s state is the Bohr radius a_B . Thus, although quantum mechanics gives a very different picture of the hydrogen atom (with the electron's probability density spread continuously through space), it agrees exactly with the Bohr model as to the electron's most probable radius in the ground state.

Armed with the radial density $P_{1s}(r)$, one can calculate several important properties of the atom. Problems 8.37 and 8.41 to 8.43 contain some examples, and here is another.

*Note that the radial density in Fig. 8.18 is zero at $r = 0$ even though the wave function itself is not. This is due to the factor r^2 in (8.84).

Example 8.4

Find the constant A in the $1s$ wave function $R_{1s} = Ae^{-r/a_B}$ and the expectation value of the potential energy for the ground state of hydrogen.

The constant A is determined by the normalization condition that the total probability of finding the electron at *any* radius must be 1.

$$\int_0^\infty P(r) dr = 1 \quad (8.86)$$

Substituting (8.85), we find that

$$4\pi A^2 \int_0^\infty r^2 e^{-2r/a_B} dr = 1 \quad (8.87)$$

This integral can be evaluated with two integrations by parts to give $a_B^3/4$ (Problem 8.41). Therefore,

$$\pi A^2 a_B^3 = 1$$

and

$$A = \frac{1}{\sqrt{\pi a_B^3}} \quad (8.88)$$

The expectation value $\langle U \rangle$ of the potential energy is the average value we would find if we measured $U(r)$ many times (always with the atom in the same state). This is found by multiplying $U(r)$ by the probability $P(r) dr$ that the electron be found at distance r , and integrating over all r .

$$\langle U \rangle = \int_0^\infty U(r)P(r) dr$$

If we substitute $U(r) = -ke^2/r$ and replace $P(r)$ by (8.85), this gives

$$\langle U \rangle = -\frac{4ke^2}{a_B^3} \int_0^\infty re^{-2r/a_B} dr$$

The integral can be evaluated by parts as $a_B^2/4$, and we find that

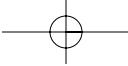
$$\langle U \rangle = -\frac{ke^2}{a_B}$$

Note that this quantum value for the mean potential energy agrees exactly with the potential energy of the Bohr model.

The 2s Wave Function

In the $n = 2$ level, with $E = -E_R/4$, we have seen that there are four independent wave functions to consider. Of these, the $2s$ wave function depends only on r

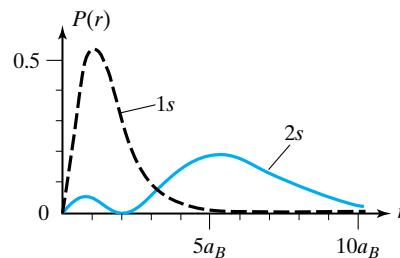
$$\psi_{2s}(r, \theta, \phi) = R_{2s}(r)$$



276 Chapter 8 • The Three-Dimensional Schrödinger Equation

FIGURE 8.19

The radial distribution $P(r)$ for the $2s$ state (solid curve). The most probable radius is $r \approx 5.2a_B$, with a small secondary maximum at $r \approx 0.76a_B$. For comparison, the dashed curve shows the $1s$ distribution on the same scale.



where $R_{2s}(r)$ is determined by the radial equation (8.79) to be (Problem 8.44)*

$$R_{2s}(r) = A \left(2 - \frac{r}{a_B} \right) e^{-r/2a_B} \quad (8.89)$$

The probability of finding the electron between distances r and $r + dr$ from the origin is again given by $P(r) dr$ with

$$P_{2s}(r) = 4\pi r^2 |R_{2s}(r)|^2 \quad (8.90)$$

This function is plotted in Fig. 8.19. As we would expect, it is peaked at a much larger radius than the $1s$ function. Specifically, the most probable radius for the $2s$ state is $r = 5.2a_B$, in approximate (though not exact) agreement with the second Bohr radius, $r = 4a_B$. An important feature of the $2s$ distribution is the small secondary maximum much closer to $r = 0$ (at $r = 0.76a_B$). This means that there is a small (but not negligible) probability of finding the $2s$ electron close to the nucleus.

The $2p_z$ Wave Function

There are three $2p$ wave functions, corresponding to the three possible orientations of an $l = 1$ state. In Section 8.6 (Example 8.3) we saw that the angular part of the $m = 0$ wave function is $\Theta(\theta) = \cos \theta$, and the complete wave function is therefore

$$\psi(r, \theta, \phi) = R_{2p}(r) \cos \theta \quad (8.91)$$

For reasons that we will see in a moment, this is often called the $2p_z$ wave function. The radial function $R_{2p}(r)$ is found by solving the radial equation for $E = -E_R/4$ and $l = 1$. This gives (Problem 8.45)

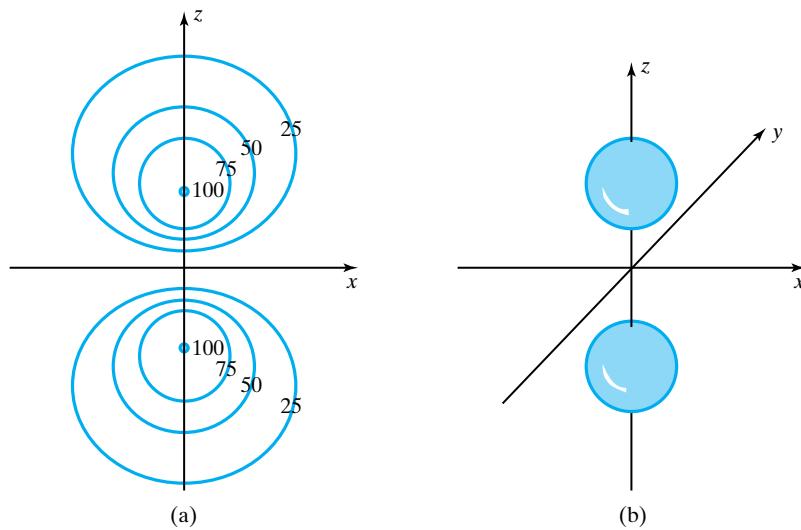
$$R_{2p}(r) = A r e^{-r/2a_B} \quad (8.92)$$

Notice that $R_{2p}(r)$ is zero at $r = 0$. Thus the probability density $|\psi|^2$ is zero at the origin, a result that applies (as already mentioned) to any state with nonzero angular momentum.

Substituting (8.92) into (8.91), we find for the complete wave function of the $2p$ state with $m = 0$.

$$\psi(r, \theta, \phi) = A r e^{-r/2a_B} \cos \theta \quad [n = 2, l = 1, m = 0] \quad (8.93)$$

*As usual, A denotes a constant, which is determined by the normalization condition (8.86). For simplicity, we use the same symbol, A , for all such constants, but we do not want to imply that they have the same value for all wave functions.

**FIGURE 8.20**

(a) Contour map of $|\psi|^2$ in the xz plane for the $2p(m = 0)$ state. The density is maximum at the points $z = \pm 2a_B$, on the z axis and zero in the xy plane. The contours shown are for $|\psi|^2$ equal to 75%, 50%, and 25% of its maximum value. **(b)** A three-dimensional view of the 75% contour, obtained by rotating the 75% contour of (a) about the z axis.

Since this depends on r and θ , it is harder to visualize than the $l = 0$ wave functions (which depend on r only). One way to show its main features is to draw a contour map of the probability density $|\psi|^2$ in the xz plane, as shown in Fig. 8.20(a). Since $|\psi|^2$ is independent of ϕ , one would find the same picture in any other plane containing the z axis, and one obtains the full three-dimensional distribution by simply rotating Fig. 8.20(a) about the z axis. Figure 8.20(b) shows a perspective view of the 75% contour obtained in this way.

The probability density $|\psi|^2$ for (8.93) is largest on the z axis (where $\cos \theta = \pm 1$) at the points $z = \pm 2a_B$ and is zero in the xy plane (where $\cos \theta = 0$). The region in which the electron is most likely to be found consists of two approximately spherical volumes centered on the z axis, one above and the other below the xy plane, as shown in Fig. 8.20(b). It is because the electron is concentrated near the z axis that the $2p$ state with $m = 0$ is called the $2p_z$ state.

The $2p_x$ and $2p_y$ Wave Functions

There are still two more $2p$ states to be discussed. An easy way to write these down is to note that the $2p_z$ wave function (8.93) can be rewritten as

$$\psi_{2p_z} = Are^{-r/2a_B} \cos \theta = Az e^{-r/2a_B} \quad (8.94)$$

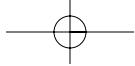
since $r \cos \theta = z$. The Schrödinger equation, from which this was derived, involves each of the coordinates x, y, z in exactly the same way. Thus, if (8.94) is a solution, so must be the two functions obtained from (8.94) by replacing z with x or with y :

$$\psi_{2p_x} = Axe^{-r/2a_B} \quad (8.95)$$

and

$$\psi_{2p_y} = Aye^{-r/2a_B} \quad (8.96)$$

The properties of these two wave functions are very similar to those of ψ_{2p_z} except that where ψ_{2p_z} is concentrated near the z axis, ψ_{2p_x} is concentrated



278 Chapter 8 • The Three-Dimensional Schrödinger Equation

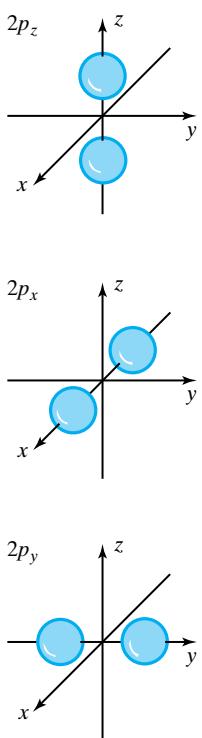


FIGURE 8.21

Perspective views of the 75% contours of $|\psi|^2$ for the $2p_z$, $2p_x$, and $2p_y$ wave functions.

near the x axis and ψ_{2p_y} near the y axis. Figure 8.21 shows perspective views of all three wave functions. We will see in Chapter 12 that the concentration of the electron near one of the axes in each of these states has important implications for the shape of some molecules.

As with s waves, it is often important to know the probability of finding the electron at a certain distance from the origin (as opposed to that for finding it at one particular position). Because the $2p$ wave functions depend on θ and ϕ as well as r , the probability of finding the electron between r and $r + dr$ must be calculated by integrating over the angles θ and ϕ . However, the result is exactly the same as (8.84) for s waves

$$P(\text{between } r \text{ and } r + dr) = P(r) dr$$

where, for any of the $2p$ states,

$$P_{2p}(r) = 4\pi r^2 |R_{2p}(r)|^2 = 4\pi A^2 r^4 e^{-r/a_B} \quad (8.97)$$

This function is plotted in Fig. 8.22, where we see that $P_{2p}(r)$ is maximum at $r = 4a_B$ (Problem 8.50); that is, the most probable radius for the $2p$ states agrees exactly with the radius of the second circular Bohr orbit.

Before leaving the $2p$ wave functions, we should mention a final complication. In our general discussion of the central-force problem, we saw that for any p state ($l = 1$) there must be three possible orientations given by $m = 1, 0$, or -1 . In the case of the $2p$ states we found, explicitly, three independent wave functions $2p_x$, $2p_y$, and $2p_z$. It turns out that these latter three wave functions are not exactly the same as the former. Specifically, the $2p_z$ state is precisely the $m = 0$ state. (This was how we derived it.) On the other hand, the $2p_x$ state is *not* the $m = 1$ nor the $m = -1$ state. Instead, the $2p_x$ wave function is the sum of the wave functions for $m = \pm 1$, while the $2p_y$ function is their difference (Problem 8.47).

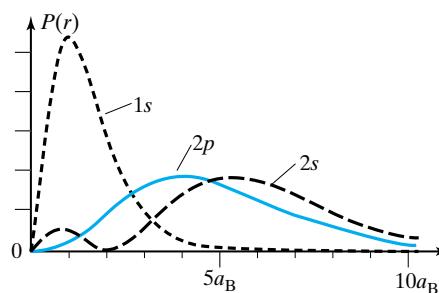
The important property of the $2p$ states is this: *Any* $2p$ wave function can be written as a combination of the three wave functions with $m = 1, 0$, and -1 , or as a combination of the wave functions $2p_x$, $2p_y$, and $2p_z$. Which set of three functions we choose to focus on is largely a matter of convenience, and for our purposes, the three functions $2p_x$, $2p_y$, and $2p_z$ are usually more suitable.

This situation is very similar to what we saw in Sections 7.6 and 7.7 when solving the differential equation $\psi'' = -k^2\psi$. Any solution of that equation could be expressed as a linear combination,

$$A \sin kx + B \cos kx$$

FIGURE 8.22

The radial probability density for the $2p$ states (solid curve). The most probable radius is $r = 4a_B$. For comparison, the dashed curves show the $1s$ and $2s$ distributions to the same scale.



of the two solutions $\sin kx$ and $\cos kx$, or as a combination,

$$Ce^{ikx} + De^{-ikx}$$

of the two solutions $e^{\pm ikx}$. When seeking the energy levels of a rigid box, we found it convenient to use the pair $\sin kx$ and $\cos kx$ to apply the boundary conditions $\psi(0) = \psi(a) = 0$. On the other hand, to interpret the solutions in terms of momentum, it was convenient to reexpress them in terms of the pair $e^{\pm ikx}$, as in Eqs. (7.77) and (7.78).

The wave function for the general state with quantum numbers n , l , and m has the form

$$\psi_{nlm}(r, \theta, \phi) = R_{nl}(r)\Theta_{lm}(\theta)e^{im\phi} \quad (8.98)$$

The first few of the angular functions $\Theta_{lm}(\theta)$ (the associated Legendre functions) were listed in Table 8.1. The first few of the radial functions $R_{lm}(r)$ are given in Table 8.2.

TABLE 8.2

The first few radial functions $R_{lm}(r)$ for the hydrogen atom. The variable ρ is an abbreviation for $\rho = r/a_B$ and a stands for a_B .

	$n = 1$	$n = 2$	$n = 3$
$l = 0$	$\frac{1}{2\sqrt{a^3}}e^{-\rho}$	$\frac{1}{\sqrt{2a^3}}\left(1 - \frac{1}{2}\rho\right)e^{-\rho/2}$	$\frac{2}{\sqrt{27a^3}}\left(1 - \frac{2}{3}\rho + \frac{2}{27}\rho^2\right)e^{-\rho/3}$
$l = 1$		$\frac{1}{\sqrt{24a^3}}\rho e^{-\rho/2}$	$\frac{8}{27\sqrt{6a^3}}\left(1 - \frac{1}{6}\rho\right)\rho e^{-\rho/3}$
$l = 2$			$\frac{4}{81\sqrt{30a^3}}\rho^2 e^{-\rho/3}$

8.9 Shells

We have seen that the most probable radius for the $1s$ state of hydrogen is $r = a_B$, while those for the $2s$ and $2p$ states are $r = 5.2a_B$ and $r = 4a_B$. For the $3s$, $3p$, and $3d$ states the most probable radii are $13.1a_B$, $12a_B$, and $9a_B$, respectively. These results are illustrated in Fig. 8.23, which shows the radial

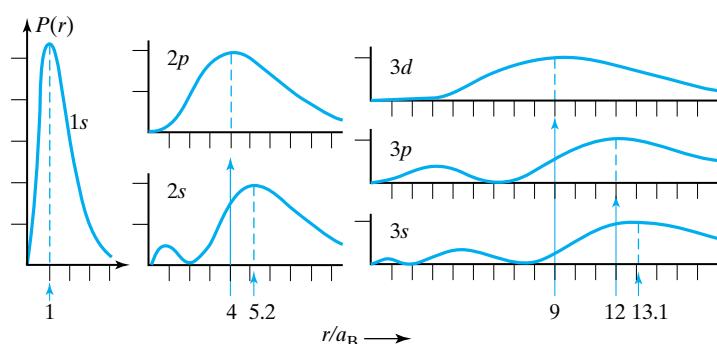
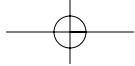


FIGURE 8.23

The radial distributions for the $n = 1, 2$, and 3 states in hydrogen. The numbers shown are the most probable radii in units of a_B .



280 Chapter 8 • The Three-Dimensional Schrödinger Equation

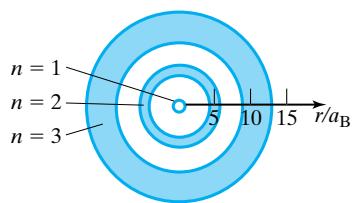


FIGURE 8.24

The most probable radius for any $n = 3$ state in hydrogen is between $9a_B$ and $13.1a_B$. The corresponding range for the $n = 2$ states is from $4a_B$ to $5.2a_B$. The most probable radius for the $n = 1$ state is a_B . These numbers define the spatial shells within which an electron with quantum number n is most likely to be found.

densities and most probable radii for all of the states concerned.

Figure 8.23 suggests what is found to be true for all of the states with which we will be concerned: For all the different states with a given value of n , the most probable radii are quite close to one another and are reasonably well separated from those with any other value of n . This important property is illustrated in a different way in Fig. 8.24, which shows how the most probable radii for the states with quantum number n are all reasonably close to the Bohr values n^2a_B , so that all electron distributions with a given n peak in the same spherical shell with radius about n^2a_B . For this reason, the word **shell** is often used for the set of all states with a given value of n .

In the hydrogen atom one can characterize a shell in two different ways that are exactly equivalent. As we have just seen, for all states in the n th shell, the most probable distances of the electron from the nucleus are clustered close to the Bohr value n^2a_B . Alternatively, since all states of a given shell have the same value of n , they all have the same energy. Thus the word “shell” can refer either to a clustering in space (what we could call a **spatial shell**) or to a clustering in energy (an **energy shell**).

The notion of shells is very important in atoms with more than one electron, as we will see in Chapter 10. We will find that the possible states of any one electron in a multielectron atom can be identified by the same three quantum numbers, n, l, m , that label the states of hydrogen. Furthermore, just as with hydrogen, all states with a given n have radial distributions that peak at about the same radius, and this most probable radius is well separated from the most probable radius for any other value of n . Thus we can speak of spatial shells, as a characteristic clustering of the radial distributions for given n , in just the same sense as in hydrogen.

On the other hand, the allowed energies of any one electron in a multielectron atom are more complicated than those of hydrogen. In particular, we will find that states with the same principal quantum number, n , do not necessarily have the same energy. Nevertheless, the states can be grouped into energy shells, such that all levels within one shell are closer to one another than to any level in a neighboring shell. However, these energy shells do not correspond to unique values of n : States with the same n may belong to different shells, and one shell may contain states with different values of n . For example, in many atoms the $3s$ and $3p$ levels are close to one another but are quite well separated from the $3d$ level, which is closer to the $4s$ and $4p$ levels. In this case the $3s$ and $3p$ levels form one energy shell, and the $3d, 4s$, and $4p$ another.

Unfortunately, the word “shell” is commonly used (without qualification) to denote both what we have called a spatial shell and what we have called an energy shell. We will discuss all this in more detail in Chapter 10. We mention it here only to emphasize that the simple situation in hydrogen (for which the grouping of states according to energy is exactly the same as the grouping according to distance from the nucleus) is unique to hydrogen.

8.10 Hydrogen-Like Ions

In Chapter 5 we saw that Bohr’s model of the hydrogen atom could be easily generalized to any hydrogen-like ion (that is, a single electron bound to a nucleus of charge Ze). The modern Schrödinger theory of hydrogen can be generalized in exactly the same way. The potential energy of the electron in hydrogen is $U = -ke^2/r$, that of the electron in a hydrogen-like ion is

$U = -Zke^2/r$. Thus the Schrödinger equation for the latter case differs from that of the former only in that ke^2 is replaced by Zke^2 in the potential-energy function.* Therefore, we can convert our hydrogen solutions into solutions for the hydrogen-like ion simply by substituting Zke^2 whenever the term ke^2 appears. This lets us draw three important conclusions with almost no additional labor.

First, the properties of the angular wave functions and the allowed values of angular momentum do not involve the potential energy U at all. Therefore, these angular properties are exactly the same for any hydrogen-like ion as for hydrogen itself.

Second, the Schrödinger equation for hydrogen has acceptable solutions only for the allowed energies,

$$E = -\frac{m_e(ke^2)^2}{2\hbar^2} \frac{1}{n^2} = -\frac{E_R}{n^2}$$

Replacing ke^2 by Zke^2 , we find for the allowed energies of a hydrogen-like ion:

$$E = -\frac{m_e(Zke^2)^2}{2\hbar^2} \frac{1}{n^2} = -Z^2 \frac{E_R}{n^2} \quad (8.99)$$

Third, the spatial extent of the hydrogen wave functions is determined by the Bohr radius,

$$\frac{\hbar^2}{m_e ke^2} = a_B$$

thus the corresponding parameter for a hydrogen-like ion is

$$\frac{\hbar^2}{m_e Zke^2} = \frac{a_B}{Z} \quad (8.100)$$

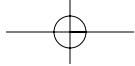
For example, the ground-state wave function of hydrogen is $\psi_{1s} = A \exp(-r/a_B)$; therefore, that for a hydrogen-like ion, with a_B replaced by a_B/Z , is

$$\psi_{1s} = Ae^{-Zr/a_B}$$

Since all wave functions are modified in the same way, each state of a hydrogen-like ion is pulled inward by a factor $1/Z$, compared to the corresponding state in hydrogen.

The relationship between the quantum properties of the hydrogen atom and the hydrogen-like ion is closely analogous to the corresponding relationship for the Bohr model. This provides the ultimate justification for the several properties of hydrogen-like ions described in Chapter 5 in connection with the Bohr model.

*Throughout this chapter we are ignoring motion of the nucleus. If we were to include this, there would be a second difference between the hydrogen atom and the hydrogen-like ion because of the different masses of the nuclei. This very small effect can be allowed for by introducing a reduced mass, as described briefly in Section 5.8, but we will ignore it here.



282 Chapter 8 • The Three-Dimensional Schrödinger Equation

When we discuss multielectron atoms in Chapter 10, we will make extensive use of the two results (8.99) and (8.100). These are so important, let us close by reiterating them in words: When an electron moves around a total charge Ze , its allowed energies are Z^2 times the allowed energies of a hydrogen atom; and its spatial distribution is scaled inward by a factor of $1/Z$ compared to hydrogen.

CHECKLIST FOR CHAPTER 8

CONCEPT	DETAILS
Partial derivatives	Derivatives with respect to one variable holding all others fixed (Sec. 8.2)
Three-dimensional Schrödinger equation	$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{2M}{\hbar^2}[U - E]\psi \quad (8.2)$
Two-dimensional rigid square box, or well	Energies, $E = E_0(n_x^2 + n_y^2)$, identified by two quantum numbers n_x and n_y (8.29)
Separation of variables	Conversion of a partial differential equation into two or more ordinary differential equations. (Secs. 8.3–8.5).
Quantum numbers	Integers or half-integers that identify the allowed values of some dynamical variable, such as energy or angular momentum
Degeneracy	Two or more physically independent wave functions with the same energy
Central-force problem	Particle subject to a force directed toward a fixed center
Polar coordinates	
Two-dimensional (r, ϕ)	Fig. 8.7
Three-dimensional (r, θ, ϕ)	Fig. 8.11
Quantization of angular momentum	
magnitude — quantum number l	$L = \sqrt{l(l+1)}\hbar, (l = 0, 1, 2, \dots) \quad (8.61)$
z component — quantum number m	$L_z = m\hbar, (m = l, l-1, \dots, -l) \quad (8.63)$
vector model	classical picture of quantum angular momenta (Sec. 8.6)
code letters, s, p, d, f, \dots	denote $l = 0, 1, 2, 3, \dots$
Hydrogen atom	
principal quantum number, n	$E_n = -E_R/n^2 \quad (8.74)$
wave functions	$\psi_{nlm}(r, \theta, \phi) = R_{nl}(r)\Theta_{lm}(\theta)e^{im\phi}$ (Tables 8.1 and 8.2)
Radial probability density $P(r)$	$P(r) dr$ = probability of finding particle between r and $r + dr \quad (8.83)$
Shells	
energy shell	group of levels with approximately the same energy
spatial shell	group of levels concentrated at approximately the same radius
Hydrogen-like ion	A single electron in the field of a charge Ze
energy levels	$E_n = -Z^2 E_R/n^2 \quad (8.99)$

PROBLEMS FOR CHAPTER 8

SECTION 8.2 (The Three-Dimensional Schrödinger Equation and Partial Derivatives)

- 8.1** • Find the two partial derivatives of (a) $x^2y^3 + x^4y^2$, (b) $(x + y)^3$, (c) $\sin x \cos y$.
- 8.2** • Find all three partial derivatives of (a) $x^2 + y^2 + z^2$, (b) $(\sin y + \cos z)^2$, (c) $x^2e^y \sin z$.
- 8.3** • (a) Supposing that $\psi(x, y, z) = f(x) + g(y) + h(z)$, find the three partial derivatives of ψ (in terms of derivatives of f , g , and h). (b) Do the same for $\psi(x, y, z) = f(x)g(y)h(z)$.
- 8.4** •• Write down the Schrödinger equation (8.2) for a free particle, subject to no forces and hence with $U(\mathbf{r}) = 0$ everywhere, and show that the function $\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$ is a solution for any fixed vector \mathbf{k} satisfying $E = \hbar^2 k^2 / 2M$. Can you suggest an interpretation for the vector \mathbf{k} ?
- 8.5** • A mountain can be described by the function $h(x, y)$, which gives the height above sea level of a point that is x east and y north of the origin O . (a) Describe in words the meaning of $\partial h/\partial x$ and $\partial h/\partial y$. (b) What does it mean to a hiker who is walking due north if $\partial h/\partial y$ is positive? (c) What if he is walking due north, but $\partial h/\partial y$ is zero and $\partial h/\partial x$ is positive?
- 8.6** •• Let $h(x, y)$ describe a mountain as in Problem 8.5. If the same mountain is given by the contour map in Fig. 8.25, give estimates for $\partial h/\partial x$ and $\partial h/\partial y$ at points P, Q, R , and the summit S . The scale for x and y (shown by the ruled line) and contours are given in meters.

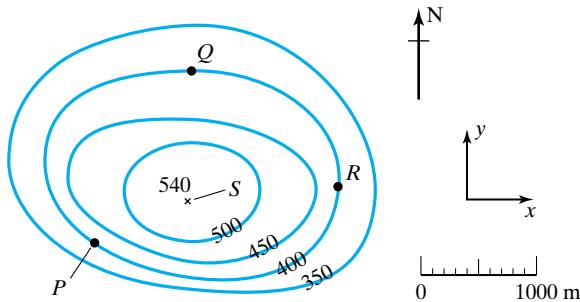


FIGURE 8.25

(Problem 8.6)

- 8.7** •• If one differentiates a function $f(x, y)$ with respect to x and then differentiates the result with respect to y , one obtains the mixed second derivative

$$\frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial y \partial x}$$

It is a theorem that for any “reasonable” function (which includes any function normally encountered in physics) it makes no difference which differentiation is done first. That is,

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} \quad (8.101)$$

A proof of this useful result is beyond the level of this book, but you can check the truth of (8.101) for some specific functions. Evaluate both sides of (8.101) for the following functions and verify that they are equal:
(a) $f = (x + y^2)^3$, **(b)** $xe^{(x+y)^2}$, **(c)** $(x + y) \ln(x - y)$.

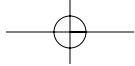
SECTION 8.3 (The Two-Dimensional Square Box)

- 8.8** • Make an energy-level diagram similar to Fig. 8.2 showing the quantum numbers, energies, and degeneracies of the lowest eight levels for a particle in a two-dimensional, rigid square box.
- 8.9** •• Consider a particle of mass M in a two-dimensional, rigid rectangular box with sides a and b . Using the method of separation of variables, find the allowed energies and wave functions for this particle. In particular, show that the allowed energies are identified by two integers n_x and n_y and have the form

$$E_{n_x, n_y} = \frac{\hbar^2 \pi^2}{2M} \left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} \right) \quad (8.102)$$

Your analysis will be very similar to that given in Section 8.3; the main purposes of this problem are that you go through that analysis yourself and understand how it generalizes to the case of a rectangle with unequal sides.

- 8.10** •• Consider a particle in a rigid rectangular box with sides a and $b = a/2$. Using the result (8.102) (Problem 8.9), find the lowest six energy levels with their quantum numbers and degeneracies.
- 8.11** •• The energy levels for a rectangular rigid box with sides a and b are given by Eq. (8.102) in Problem 8.9. When $a \neq b$, some of the degeneracies noticed for the square box (Fig. 8.2) are no longer present. To illustrate this, find the lowest six levels and their degeneracies for the case $a = 1.1b$. Compare with the levels for the case $a = b$ (Fig. 8.2). Your results will illustrate a general trend: When one reduces the symmetry of a system, its degeneracies usually decrease.
- 8.12** •• (a) Consider the state with $n_x = 1$ and $n_y = 2$ for a particle in a two-dimensional, rigid square box. Write down $|\psi|^2$. At what points is the particle most likely to be found? How many such points are there? Sketch a contour map similar to those in Fig. 8.4. (b) Repeat for $n_x = 2, n_y = 2$. (c) Repeat for $n_x = 4, n_y = 3$.
- 8.13** •• The energy levels of a particle in a cubical box can be found from Eq. (8.103) (Problem 8.15) by setting $a = b = c$. Find the lowest eight energy levels for a particle in a three-dimensional, rigid cubical box. Draw an energy-level diagram for these levels, showing their quantum numbers, energies, and degeneracies.
- 8.14** •• In Chapter 7 we claimed that an electron confined inside a thin conducting wire was essentially a one-dimensional system. To illustrate this, take as a model of the wire a long thin rigid box of length a and



284 Chapter 8 • The Three-Dimensional Schrödinger Equation

square cross section $b \times b$ (with $a \gg b$). (a) Using the formula (8.103) (Problem 8.15), write down the ground-state energy for an electron in this box. (b) Write down the energy, measured up from the ground state, of the general excited state. (c) Do the same for an electron in a one-dimensional box of the same length a . (d) Suppose that $a = 1\text{ m}$ and $b = 1\text{ mm}$. Show that the first 1700 (approximately) levels of the electron in the wire are identical to those for the one-dimensional box.

- 8.15** ••• Show that the allowed energies of a mass M confined in a three-dimensional rectangular rigid box with sides a, b , and c are

$$E = \frac{\hbar^2 \pi^2}{2M} \left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right) \quad (8.103)$$

where the three quantum numbers n_x, n_y, n_z are any three positive integers $(1, 2, 3, \dots)$. [Hint: Use separation of variables, and seek a solution of the form $\psi = X(x)Y(y)Z(z)$. Note that by setting $a = b = c$, one obtains the cubical box of Example 8.2.]

SECTION 8.4 (The Two-Dimensional Central-Force Problem)

- 8.16** • (a) For the two-dimensional polar coordinates defined in Fig. 8.7 (Section 8.4), prove the relations

$$x = r \cos \phi \quad \text{and} \quad y = r \sin \phi \quad (8.104)$$

(b) Find corresponding expressions for r and ϕ in terms of x and y .

- 8.17** • A certain point P in two dimensions has rectangular coordinates (x, y) and polar coordinates (r, ϕ) . What are the polar coordinates of the point $Q(-x, -y)$? Illustrate your answer with a picture.

- 8.18** •• Changes of coordinates in two dimensions (such as that from x, y to r, ϕ) are much more complicated than in one dimension. In one dimension, if we have a function $f(x)$ and choose to regard x as a function of some other variable u , then the derivative of f with respect to u is given by the chain rule,

$$\frac{df}{du} = \frac{df}{dx} \frac{dx}{du}$$

In two dimensions the chain rule reads

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial r} \quad (8.105)$$

and

$$\frac{\partial f}{\partial \phi} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial \phi} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial \phi}$$

- (a) Use the relations (8.104) (Problem 8.16) to evaluate the four derivatives $\partial x/\partial r$, $\partial y/\partial r$, $\partial x/\partial \phi$, and $\partial y/\partial \phi$. (b) If $f = \exp \sqrt{x^2 + y^2}$, use (8.105) to find $\partial f/\partial r$. (c) What is $\partial f/\partial \phi$? (d) By noticing that $\sqrt{x^2 + y^2} = r$ and hence that $f = \exp(r)$, evaluate $\partial f/\partial r$ and $\partial f/\partial \phi$ directly and check that your answers in parts (b) and (c) are correct.

- 8.19** ••• A crucial step in solving the Schrödinger equation for the central-force problem was the identity (8.33).

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{\partial^2 \psi}{\partial r^2} + \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \phi^2} \quad (8.106)$$

The purpose of this problem is to prove this identity by showing that the right-hand side is equal to the left.

- (a) If you have not already done so, do part (a) of Problem 8.18. (b) Use the chain rule (8.105) to show that

$$\frac{\partial \psi}{\partial r} = \frac{\partial \psi}{\partial x} \cos \phi + \frac{\partial \psi}{\partial y} \sin \phi$$

- (c) Use the chain rule on each term in $\partial \psi / \partial r$ to find $\partial^2 \psi / \partial r^2$ in terms of $\partial^2 \psi / \partial x^2$, $\partial^2 \psi / \partial x \partial y$, $\partial^2 \psi / \partial y^2$. [Recall (8.101) and remember that $\partial f / \partial r$ denotes the derivative with respect to r when ϕ is fixed; therefore, $\partial \phi / \partial r = 0$.] (d) Similarly, find $\partial^2 \psi / \partial \phi^2$ in terms of derivatives with respect to x and y . Remember that $\partial r / \partial \phi = 0$. (e) Substitute the results of the previous three parts into the right-hand side of (8.106), and show that you get the left-hand side.

SECTION 8.5 (The Three-Dimensional Central-Force Problem)

- 8.20** •• The spherical polar coordinates (r, θ, ϕ) are defined in Fig. 8.11. Derive the expressions given there for x, y , and z in terms of (r, θ, ϕ) . Find corresponding expression for r, θ , and ϕ in terms of (x, y, z) .

- 8.21** •• A point P on the earth's surface has rectangular coordinates (x, y, z) and spherical polar coordinates (r, θ, ϕ) (with coordinates defined so that the origin is at the earth's center and the z axis points north). What are the coordinates (rectangular and spherical) of the place Q at the opposite end of the earth diameter through P ?

- 8.22** •• Substitute the separated form $\psi = R(r)\Theta(\theta)\Phi(\phi)$ into the Schrödinger equation (8.49). (a) Show that if you multiply through by $r^2 \sin^2 \theta / (R\Theta\Phi)$ and rearrange, you get an equation of the form $\Phi''/\Phi =$ (function of r and θ). Explain clearly why each side of this equation must be a constant, which we can call $-m^2$. (b) Show that the resulting equation, (function of r and θ) $= -m^2$, can be put in the form

$$\frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) - \frac{m^2}{\sin^2 \theta} = (\text{function of } r)$$

Explain (again) why each side of this equation must be a constant, which we can call $-k$. Derive the r and θ equations (8.54) and (8.53).

SECTION 8.6 (Quantization of Angular Momentum)

- 8.23** • Consider the vector model for the case $l = 2$. Referring to Fig. 8.14, find the minimum possible angle between \mathbf{L} and the z axis.

- 8.24** • (a) Draw a vector model diagram similar to Fig. 8.14 for angular momentum of magnitude given by $l = 1$. (b) How many possible orientations are there? (c) What is the minimum angle between \mathbf{L} and the z axis?

- 8.25** • Do the same tasks as in Problem 8.24 but for $l = 3$.
- 8.26** • For a given magnitude $L = \sqrt{l(l+1)\hbar}$ of \mathbf{L} , what is the largest allowed value of L_z ? Prove that this largest value of L_z is less than or equal to L (as one would certainly expect classically).
- 8.27** • The allowed magnitudes of the angular momentum are $L = \sqrt{l(l+1)\hbar}$, whereas the Bohr model assumes that $L = l\hbar$. (In both cases, l is restricted to integers.) Compute the ratio $L(\text{correct})/L(\text{Bohr})$ for $l = 1, 2, 3, 4, 10$, and 100. Comment.
- 8.28** •• The allowed magnitudes of angular momentum are $L = \sqrt{l(l+1)\hbar}$. Use the binomial expansion to prove that when l is large, $L \approx (l + \frac{1}{2})\hbar$. (This shows that even for large l , modern quantum mechanics does not quite agree with the Bohr model.)
- 8.29** •• Write down the θ equation (8.65) for the special case that $l = m = 0$. (a) Verify that $\Theta = \text{constant}$ is a solution. (b) Verify that a second solution is $\Theta = \ln[(1 + \cos \theta)/(1 - \cos \theta)]$, and show that this is infinite when $\theta = 0$ or π (and hence is unacceptable). (c) Since the θ equation is a second-order differential equation, any solution must be a linear combination of these two. Write down the general solution, and prove that the only acceptable solution is $\Theta = \text{constant}$.
- 8.30** •• Write down the θ equation (8.65) for the case $l = m = 1$. Verify that $\Theta = \sin \theta$ is a solution. (Any other solution is infinite at $\theta = 0$ or π , so $\sin \theta$ is the only acceptable solution.) Write down the complete wave function (8.64), showing its explicit dependence on θ and ϕ for $l = 1, m = 1$ and for $l = 1, m = -1$. [You don't know the radial function, so just leave it as $R(r)$.] With r fixed, in what directions is $|\psi|^2$ a maximum for these states?
- 8.31** •• Explain clearly why the angular function $\Theta_{lm}(\theta)$ must satisfy $\Theta_{l,-m}(\theta) \propto \Theta_{l,m}(\theta)$. [Hint: We have stated (though not actually proved) that the θ equation (8.65) has, at most, one independent acceptable solution for any given values of l and m .]
- 8.32** •• Make a table of all the spherical harmonics, as defined in (8.69), for $l = 0, 1, 2$ and for all corresponding values of m .
- 8.33** •• The normalization condition for a three-dimensional wave function is $\int |\psi|^2 dV = 1$. (a) Show that in spherical polar coordinates, the element of volume is $dV = r^2 dr \sin \theta d\theta d\phi$. [Hint: Think about the infinitesimal volume between r and $r + dr$, between θ and $\theta + d\theta$, and between ϕ and $\phi + d\phi$.] (b) Show that if $\psi = R(r)Y(\theta, \phi)$, the normalization integral is the product of two terms

$$\begin{aligned} \int |\psi|^2 dV &= \left(\int_0^\infty |R(r)|^2 r^2 dr \right) \\ &\quad \left(\int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi |Y(\theta, \phi)|^2 \right) = 1 \end{aligned}$$

(c) It is usually convenient to normalize the functions $R(r)$ and $Y(\theta, \phi)$ separately, so that each of the factors in this middle expression is equal to 1. Verify that all of the spherical harmonics $Y_{lm}(\theta, \phi)$ with $l = 0$ or 1 do satisfy

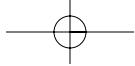
$$\int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi |Y_{lm}(\theta, \phi)|^2 = 1$$

The required spherical harmonics are defined in (8.69) and Table 8.1.

- 8.34** ••• If you haven't already done so, do parts (a) and (b) of Problem 8.33, and then do part (c), but for the five spherical harmonics with $l = 2$.

SECTION 8.7 and 8.8 (The Energy Levels of the Hydrogen Atom and Hydrogenic Wave Functions)

- 8.35** • Prove that the degeneracy of the n th level in the hydrogen atom is n^2 ; that is, verify the result (8.77). (But be aware that this number gets doubled because of the electron's spin, as we describe in Chapter 9.)
- 8.36** • It is known that a certain hydrogen atom has a definite value of l . (a) What does this statement tell you about the angular momentum? (b) What are the allowed energies consistent with this information?
- 8.37** • The mean value (or expectation value) of $1/r$ for any state is $\langle 1/r \rangle = \int_0^\infty (1/r)P(r) dr$. Find $\langle 1/r \rangle$ for the $1s$ state of hydrogen. Comment. [Hint: See the integrals in Appendix B.]
- 8.38** •• (a) It is known that a certain hydrogen atom has $n = 5$ and $m = 2$. How many different states are consistent with this information? (b) Answer the same question (in terms of n and m) for arbitrary values of n and m .
- 8.39** •• (a) The radial equation for $l = 0$ states in hydrogen was given in (8.79). (a) Verify that this can be rewritten as
- $$\frac{d^2}{dr^2}(rR) = \left(\frac{1}{n^2 a_B^2} - \frac{2}{a_B r} \right)(rR) \quad (8.107)$$
- (b) For the case that $n = 1$, prove that $R_{1s} = e^{-r/a_B}$ is a solution of this equation (that is, calculate the derivative on the left and show that it is equal to the right-hand side).
- 8.40** •• The hydrogenic radial functions $R(r)$ are relatively simple for the case $l = n - 1$ (the maximum allowed value of l for given n).
- $$R(r) = Ar^{n-1}e^{-r/a_B} \quad [l = n - 1] \quad (8.108)$$
- (a) Write down the radial Schrödinger equation, (8.72), for this case. (b) Verify that the proposed solution (8.108) does indeed satisfy this equation if and only if $E = -E_R/n^2$.
- 8.41** •• Use integration by parts to evaluate the integral in (8.87), and hence verify that the normalization constant for the $1s$ wave function is $A = 1/\sqrt{\pi a_B^3}$.



286 Chapter 8 • The Three-Dimensional Schrödinger Equation

8.42 •• The average (or expectation) value $\langle r \rangle$ of the radius for any state is $\int_0^\infty r P(r) dr$. Find $\langle r \rangle$ for the $1s$ state of hydrogen. Referring to Fig. 8.18, explain the difference between the average and most probable radii.

8.43 ••• The probability of finding the electron in the region $r > a$ is $\int_a^\infty P(r) dr$. What is the probability that a $1s$ electron in hydrogen would be found outside the Bohr radius ($r > a_B$)?

8.44 ••• (a) Write down the radial equation (8.107) for the case that $n = 2$ and $l = 0$ and verify that

$$R_{2s} = A \left(2 - \frac{r}{a_B} \right) e^{-r/2a_B}$$

is a solution. (b) Use the normalization condition (8.86) to find the constant A . (See Appendix B.)

8.45 ••• Write down the radial equation (8.72) for the case that $n = 2$ and $l = 1$. Put in the value $-E_R/4$ for the energy and use the known expressions for a_B and E_R to eliminate all dimensional constants except a_B [as was done in (8.80)]. Verify that $R_{2p} = Are^{-r/2a_B}$ is a solution, and use the normalization condition (8.86) with $P_{2p} = 4\pi r^2 |R_{2p}|^2$ to prove that $A = 1/(4\sqrt{6\pi a_B^5})$.

8.46 ••• (a) Use the wave function R_{2p} with the normalization constant A as given in Problem 8.45 to find the average (or expectation) value of the radius, $\langle r \rangle = \int_0^\infty r P(r) dr$, for any of the $2p$ states of hydrogen. (b) Find the average potential energy. (c) Compare your results with the values predicted by the Bohr model. (Do they agree exactly? Roughly?)

8.47 ••• (a) Write down the θ equation (8.65) for the $2p$ states with $m = \pm 1$. Show that the solution is $\Theta(\theta) = \sin \theta$. (There are, of course, two solutions of this second-order equation, but this is the only acceptable one.) This means that the complete wave functions for the $2p$ states with $m = \pm 1$ are

$$\psi_{2,1,\pm 1} = R_{2p}(r) \sin \theta e^{\pm i\phi}$$

(b) Prove that the sum of these two wave functions is the $2p_x$ wave function (times an uninteresting factor of 2) and that the difference is the $2p_y$ function (times $2i$). [Hint: Rewrite $e^{\pm i\phi}$ as $\cos \theta \pm i \sin \theta$, and remember the relations for x and y in terms of r, θ, ϕ in Fig. 8.11.]

SECTION 8.9 (Shells)

8.48 • Consider the radial probability density $P(r)$ for the ground state of hydrogen, as given by Eq. (8.85). By finding where $P(r)$ is maximum, find the most probable radius for this state.

8.49 • Using the wave function (8.108) given in Problem 8.40, write down the radial probability density for a hydrogen atom in a state with $l = n - 1$. Find the most probable radius. Notice that in this case (with l equal to its maximum possible value, $l = n - 1$) the quantum mechanical answer agrees with the Bohr model.

8.50 •• Write down the radial density $P(r)$ for the $2s$ and $2p$ states of hydrogen. [See (8.90) and (8.97).] Find the most probable radius for each of these states. [Hint: If $P(r)$ is maximum, so is $\sqrt{P(r)}$.]

SECTION 8.10 (Hydrogen-Like Ions)

8.51 • What is the most probable radius for a $1s$ electron in the hydrogen-like ion Ni^{27+} ? What is its binding energy?

8.52 • An inner electron in a heavy atom is affected relatively little by the other electrons and hence has a wave function very like that for a single electron in orbit around the same nucleus. Approximately, what is the most probable radius for a $1s$ electron in lead? What is this electron's approximate binding energy?

8.53 • A hydrogen-like ion Mg^{11+} drops from its $n = 2$ to its $n = 1$ level. What is the wavelength of the photon emitted? What sort of radiation is this?

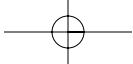
8.54 •• A hydrogen-like ion of calcium emits a photon with energy $E_\gamma = 756 \text{ eV}$. What transition is involved?

COMPUTER PROBLEMS

8.55 •• (Section 8.8) Use suitable plotting software to plot the radial probability distributions for all hydrogen states with $n = 1, 2$, and 3 . Use the same horizontal scale for all plots, and compare your results with Fig. 8.23.

8.56 •• (Section 8.10) Use suitable plotting software to plot the radial probability distributions for the $1s$ states of hydrogen and of He^+ . Put both curves on the same plot, and comment on any differences.

8.57 ••• (Section 8.6) The study (either theoretical or numerical) of the θ equation (8.65) for $\Theta(\theta)$ is made difficult because the differential equation has singularities at $\theta = 0$ and π . (The $\sin \theta$ in the denominator makes the first term infinite at $\theta = 0$ and π . Indeed this is exactly why there are no acceptable solutions for most values of l .) Nevertheless, if you have access to software that can solve differential equations numerically, you can get some insight to the acceptable and unacceptable solutions. (a) Write down the differential equation (8.65) for $m = 0$ and $l = 2$. Solve it numerically for the boundary conditions $\Theta(\pi/2) = 1$ and $\Theta'(\pi/2) = 0$. Plot your result for $0 \leq \theta \leq \pi$, and note that this solution looks perfectly acceptable. Repeat with the boundary conditions $\Theta(\pi/2) = 0$ and $\Theta'(\pi/2) = 1$ and explain why this solution appears to be unacceptable. (b) Repeat part (a) with $m = 0$ but $l = 1.75$. Explain why both solutions appear to be unacceptable; that is, there is no acceptable solution for these values of m and l .



Chapter 9

Electron Spin

- 9.1 Introduction
- 9.2 Spin Angular Momentum
- 9.3 Magnetic Moments
- 9.4 The Zeeman Effect
- 9.5 Spin Magnetic Moments
- 9.6 The Anomalous Zeeman Effect
- 9.7 Fine Structure*
- 9.8 Magnetic Resonance Imaging*

*These sections can be omitted without loss of continuity.

9.1 Introduction

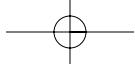
In Chapter 8 we saw how the Schrödinger equation can explain many properties of the hydrogen atom. Our next logical step would be to describe how Schrödinger's theory — unlike the Bohr model — also gives an excellent account of all higher, multielectron atoms. Before we can do this, however, we need to introduce another important property of the electron, its *spin angular momentum*, or *spin*. In this chapter we describe the electron's spin and several of its experimental consequences. Then in Chapter 10 we will return to the Schrödinger equation and use it to explain the properties of multielectron atoms.

In Section 9.2 we state the observed facts about the electron's spin angular momentum and its quantized values. Most of the more obvious manifestations of the electron's spin concern the magnetic effects associated with a spinning charged particle. Therefore, in Sections 9.3 to 9.5 we describe the magnetic properties of orbiting and spinning charged particles. Then in Sections 9.6 and 9.7 we describe several important experimental consequences of the electron's spin magnetic moment. These effects, all important in their own right, are historically important because they gave evidence for the existence of spin. Finally, in Section 9.8 we discuss an important technological application of spin: magnetic resonance imaging, a powerful medical diagnostic tool.

9.2 Spin Angular Momentum

As the earth orbits around the sun its total angular momentum \mathbf{J} is the sum of two terms,

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \quad (9.1)$$



288 Chapter 9 • Electron Spin

Here the first term, \mathbf{L} , is $\mathbf{r} \times \mathbf{p}$, where \mathbf{r} is the position vector of the earth relative to the sun and \mathbf{p} is the earth's linear momentum. Because this term arises from the earth's yearly orbital motion, it is called the *orbital angular momentum*. The second term, \mathbf{S} , is $I\omega$, where I is the earth's moment of inertia and ω is the angular velocity of its daily spinning on its own axis; this second term, \mathbf{S} , is called the earth's *spin*. In a similar way, the angular momentum of an electron is found to be the sum of two terms with the same form (9.1). The first term, \mathbf{L} , is the orbital angular momentum, and this is the angular momentum discussed in Chapter 8, with quantized magnitude $\sqrt{l(l+1)\hbar}$ and components $m\hbar$. The second term, \mathbf{S} , is called the electron's **spin**. For most purposes, one can visualize the electron's spin as analogous to the earth's spinning motion as it rotates on its own axis.*

We saw in Chapter 8 that the magnitude of \mathbf{L} is quantized, with allowed values

$$L = \sqrt{l(l+1)\hbar}$$

The magnitude of the spin \mathbf{S} is found to be given by a similar formula,

$$S = \sqrt{s(s+1)\hbar} \quad (9.2)$$

Here the spin quantum number s determines the magnitude of the spin \mathbf{S} in just the same way that l determines the magnitude of \mathbf{L} . There is, however, an important difference: As we saw in Chapter 8, l can be any integer:

$$l = 0, 1, 2, 3, \dots$$

Experiment shows that, for an electron, s always has a fixed value, which is not an integer, namely

$$s = \frac{1}{2} \quad (9.3)$$

According to (9.2), this means that the electron's spin \mathbf{S} always has the same magnitude,

$$S = \sqrt{s(s+1)\hbar} = \sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)\hbar} = \frac{\sqrt{3}}{2}\hbar \quad (9.4)$$

and for this reason, the spin is sometimes described as the *intrinsic angular momentum* of the electron. Because the quantum number s is $\frac{1}{2}$, one often refers to the electron as having "spin half."

The possible values of the z component (or any other component) of the orbital angular momentum \mathbf{L} have the form

$$L_z = m\hbar$$

where m runs in integer steps from l to $-l$:

$$m = l, l-1, \dots, -l$$

*We should emphasize that this analogy is not exact. For example, if the electron were, like the earth, a spinning ball of matter, its spin angular momentum would be characterized by a quantum number, l , that could take on any integer value, $l = 0, 1, 2, \dots$. As we will see shortly, the spin quantum number is *fixed* and *noninteger*.

Section 9.2 • Spin Angular Momentum **289**

It is found that a corresponding result holds for spin. The possible values of S_z are

$$S_z = m_s \hbar$$

where m_s is a quantum number that runs in integer steps from s to $-s$. However, since $s = \frac{1}{2}$, this gives only two possible values:

$$m_s = \frac{1}{2} \text{ or } -\frac{1}{2}$$

and hence

$$S_z = \frac{1}{2}\hbar \text{ or } -\frac{1}{2}\hbar \quad (9.5)$$

We often describe these two possibilities by saying the electron's spin can be "up" or "down," and represent these two states by arrows, \uparrow and \downarrow . However, note that in neither case is \mathbf{S} actually parallel to the z axis since S_z is smaller than S , as is clear from (9.4) and (9.5). This is the same thing that we saw with respect to \mathbf{L} : Even when L_z has its maximum value, it is smaller than L , so that \mathbf{L} is never exactly parallel to the z axis.

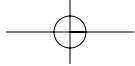
A complete specification of an electron's state of motion requires that one specify its spin orientation as well as its orbital motion. For example, for an electron in a hydrogen atom, the quantum numbers n, l, m specify the orbital motion, but for each choice of n, l, m , the spin can be either up or down, corresponding to $m_s = \frac{1}{2}$ or $-\frac{1}{2}$. It turns out that the energy of the H atom is almost completely independent of the spin orientation. Thus the allowed energies calculated in Chapter 8 are still correct, but there are twice as many independent states in each energy level as we had calculated formerly. The ground state, with $n = 1$ and $l = m = 0$, can have $m_s = \pm \frac{1}{2}$ and is therefore twofold degenerate. More generally, we saw that for the n th energy level there are n^2 possible values of l and m ; for each of these there are two possible spin orientations. Therefore, the total degeneracy of the n th level is $2n^2$:

$$\text{(degeneracy of } n\text{th level in hydrogen)} = 2n^2 \quad (9.6)$$

When an electron's state of motion is completely specified, we say that the electron is in a definite **quantum state**, or just **state**; for example, we can speak of the quantum state identified by the four quantum numbers n, l, m , and m_s in hydrogen. A specification of an electron's orbital motion, but not its spin orientation, is sometimes called an **orbital**; for example, we can speak of the orbital given by the three quantum numbers n, l , and m . For each orbital, there are evidently two independent quantum states, corresponding to the two possible values of m_s .

Most of the elementary particles have a spin angular momentum. For example, both the proton and the neutron, like the electron, have spin half; that is, they have a spin angular momentum whose magnitude is given by (9.2), with $s = \frac{1}{2}$. In fact, all elementary particles have a spin angular momentum given by (9.2), although different particles may have different values of the spin quantum number s . Thus the photon is found to have $s = 1$ ("spin one"). For the pion, $s = 0$; that is, the pion has no spin angular momentum at all. For a particle called the Δ (delta), $s = \frac{3}{2}$ (see Chapter 18), and so on.

As far as we know, the electron is entirely elementary. That is, we have no evidence for any internal constituents of the electron. In particular, the spin angular momentum of the electron is not, as far as we know, the result of the



290 Chapter 9 • Electron Spin

internal motion of any “sub-elementary” particles. To emphasize this point, we sometimes say that its spin is an *intrinsic* property of the electron. By contrast, we now know that the proton is made up of more fundamental particles called quarks and gluons, and the proton’s spin angular momentum is really just the vector sum of the angular momentum of these constituents. In a similar way, from the point of view of an atomic physicist, the atomic nucleus behaves just like a structureless “elementary” particle. The internal motions of its constituent protons and neutrons may give a nucleus some angular momentum, and this angular momentum is given by a formula just like (9.2). We often refer to this angular momentum as the nuclear “spin,” even though we know that it is just the vector sum of the angular momenta of all the constituent protons and neutrons. In this chapter our focus is almost entirely on the electron, for which the spin is an intrinsic property with $s = \frac{1}{2}$.

9.3 Magnetic Moments

There is an enormous body of evidence for the electron’s spin angular momentum. However, most of this evidence is indirect. In particular, much of the evidence for spin relates not to the angular momentum itself, but to the magnetic moment associated with any rotating electric charge. We must, therefore, review the concept of magnetic moment and describe its relation to the rotational motion of a charged particle. In this section and Section 9.4 we confine ourselves to the magnetic properties associated with the orbital motion of an electron. Then in Section 9.5 we describe the additional magnetic moment that results from the electron’s spin. Finally, in Sections 9.6 and 9.7 we describe some phenomena in which the spin magnetic moment plays an important role.

We start by considering the magnetic properties of a classical point electron traveling around a nucleus in a circular orbit. An orbiting charge acts like a small current loop, and we know from classical electromagnetic theory that a current loop both produces a magnetic field and responds to an externally applied field. If a current i flowing around a small plane loop of area A is placed in a magnetic field \mathbf{B} , it experiences a torque Γ given by

$$\Gamma = i\mathbf{A} \times \mathbf{B}$$

where the vector \mathbf{A} has magnitude equal to the area A and is perpendicular to the plane of the loop as in Fig. 9.1. The sense of the vector \mathbf{A} is given by the familiar right-hand rule: If you curl the fingers of your right hand around the loop in the direction of i , your thumb will point in the direction of \mathbf{A} .

It is usual to rewrite the torque $\Gamma = i\mathbf{A} \times \mathbf{B}$ as

$$\Gamma = \mu \times \mathbf{B} \quad (9.7)$$

where the vector μ ,

$$\mu = i\mathbf{A} \quad (9.8)$$

is called the **magnetic moment** of the loop. The torque Γ tends to turn the loop so that μ points in the same direction as the magnetic field \mathbf{B} .

Because of the torque (9.7), a current loop in a B field has a potential energy U that depends on the loop’s orientation. To evaluate this energy, we recall that the work done by a torque Γ as it turns through an angle $d\theta$ is $\Gamma d\theta$.

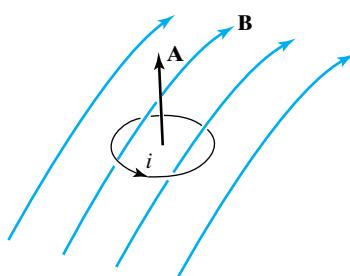


FIGURE 9.1

The current i flows around a loop specified by the vector \mathbf{A} , whose magnitude gives the loop’s area and whose direction is perpendicular to the plane of the loop. A magnetic field \mathbf{B} exerts a torque that tends to align \mathbf{A} with \mathbf{B} .

The torque (9.7) has magnitude $\mu B \sin \theta$ and is in a direction to decrease θ . Thus the work done by B when the loop is brought to angle θ is

$$W = - \int \Gamma d\theta = -\mu B \int \sin \theta d\theta = \mu B \cos \theta + \text{constant} \quad (9.9)$$

The potential energy U is defined as the negative of this work. Since the definition of potential energy always contains an arbitrary constant, it is customary to set the constant in (9.9) equal to zero, with the result that

$$U = -\mu B \cos \theta = -\boldsymbol{\mu} \cdot \mathbf{B} \quad (9.10)$$

Notice that this potential energy is minimum when $\theta = 0$, with $\boldsymbol{\mu}$ pointing along \mathbf{B} in stable equilibrium (see Fig. 9.2).

Let us now consider the current loop produced by a (still classical) electron in circular orbit with radius r and period T . The current i is equal to the total charge passing any fixed point in unit time. The electron has charge of magnitude e , and speed $v = 2\pi r/T$. Therefore, the current has magnitude

$$i = \frac{e}{T} = e \frac{v}{2\pi r}$$

and the magnetic moment has magnitude

$$\mu = iA = \frac{ev}{2\pi r} \pi r^2 = \frac{1}{2} evr \quad (9.11)$$

It is convenient to relate the magnetic moment μ to the angular momentum L . (Since μ and L both result from the electron's orbital motion, one might expect some simple relation between them.) Since the angular momentum has magnitude

$$L = m_e vr$$

(where m_e denotes the electron mass), we see from (9.11) that

$$\frac{\mu}{L} = \frac{e}{2m_e} \quad (9.12)$$

We conclude that the ratio of μ to L — the so-called **gyromagnetic ratio** — is a constant that depends only on the charge and mass of the electron. Because the electron's charge is negative, the current is in a direction opposite to the electron's velocity, so that the vectors $\boldsymbol{\mu}$ and \mathbf{L} are antiparallel. Thus, we can rewrite (9.12) in vector form as

$$\boldsymbol{\mu} = -\frac{e}{2m_e} \mathbf{L} \quad (9.13)$$

We have derived the result (9.13) for the magnetic moment of a classical point electron in a circular orbit. In quantum mechanics it turns out that exactly the same expression correctly predicts the magnetic moment due to the orbital motion of an electron, provided that we use the correct quantum values

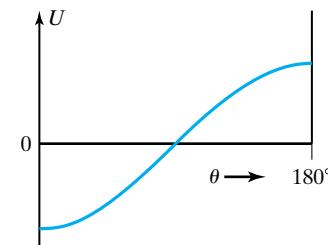
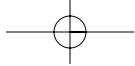


FIGURE 9.2

The potential energy $U = -\mu B \cos \theta$ of a magnetic moment $\boldsymbol{\mu}$ in a field \mathbf{B} is minimum when $\theta = 0$ and $\boldsymbol{\mu}$ is parallel to \mathbf{B} .



292 Chapter 9 • Electron Spin

for the magnitude and components of \mathbf{L} . In particular, for a given magnitude $\sqrt{l(l+1)\hbar}$ we know that \mathbf{L} has just $2l + 1$ possible orientations. According to (9.13), the same is true of the magnetic moment μ of an orbiting electron: For a given value of l , μ has just $2l + 1$ possible orientations.

Equation (9.13) gives the magnetic moment due to the orbital motion of an electron. As one might expect, there is an additional magnetic moment due to the electron's spinning motion. Before we discuss this spin magnetic moment, we use (9.10) and (9.13) to see how the energy of an atom changes when it is put in a magnetic field.

9.4 The Zeeman Effect

Because of the motion of their electrons, most atoms have a magnetic moment μ . Therefore, by applying a magnetic field \mathbf{B} , one can change an atom's energy levels by an amount $-\mu \cdot \mathbf{B}$ as in (9.10). This means that the energies of photons emitted and absorbed by the atom will change. That is, by putting an atom in a magnetic field, one can change its spectrum. This effect was first observed in 1896 by the Dutch physicist Pieter Zeeman and is called the **Zeeman effect**.

To simplify our discussion, we consider at first an atom in which the magnetic moments due to the electrons' spins cancel out. The simplest atom in which this can happen is helium, with its two electrons, and this is the atom that we consider. In many states of helium, including the ground state, the spins of the two electrons point in opposite directions so that the total magnetic moment due to the spins is zero. These states with zero total spin are called **singlet states**. Furthermore, it turns out that in all the states of helium, one of the electrons has zero orbital angular momentum. Therefore, the total magnetic moment for any of the singlet states of helium is just the moment $\mu = -(e/2m_e)\mathbf{L}$ due to the orbital motion of the second electron.

In the absence of a magnetic field, the helium atom has an energy that we call E_0 , and its angular momentum (the orbital momentum \mathbf{L} of the second electron) has a magnitude given by the quantum number $l = 0, 1, 2, \dots$. The angular momentum \mathbf{L} has $2l + 1$ possible different orientations, corresponding to the $2l + 1$ possible values of $L_z = m\hbar$, with $m = l, l - 1, \dots, -l$. In the absence of a magnetic field, the energy is the same for all of these states, and the level E_0 is $(2l + 1)$ -fold degenerate.

Suppose now that we apply a magnetic field \mathbf{B} to our atom. According to (9.10) this will change the atom's energy by the amount $-\mu \cdot \mathbf{B}$, which depends on the orientation of μ . Now μ is given by (9.13) and has $2l + 1$ different possible orientations. Therefore, we can anticipate *not only that the energy will change as a result of the magnetic field, but that it will change by a different amount for each of the $2l + 1$ different orientations*. That is, by applying a magnetic field, we remove the $(2l + 1)$ -fold degeneracy of the original energy level.

The size of the energy shift due to the magnetic field is easily calculated: With the field switched on, we denote the total energy by $E = E_0 + \Delta E$, where the shift ΔE is given by (9.10) and (9.13) as

$$\Delta E = -\mu \cdot \mathbf{B} \quad (9.14)$$

$$= \left(\frac{e}{2m_e} \right) \mathbf{L} \cdot \mathbf{B} \quad (9.15)$$

Pieter Zeeman
(1865–1943, Dutch)



At the suggestion of his teacher Lorentz, Zeeman investigated the effect of magnetic fields on atomic spectra. The results confirmed Lorentz's suspicion that atomic spectra are somehow connected to the motion of electrons in the atoms. Zeeman and Lorentz shared the 1902 Nobel Prize in physics for this work.

If we choose our z axis in the direction of the applied field \mathbf{B} , then (9.15) simplifies to

$$\Delta E = \left(\frac{e}{2m_e} \right) L_z B$$

or, since the possible values of L_z are $m\hbar$,

$$\Delta E = \left(\frac{e\hbar}{2m_e} \right) m B \quad (9.16)$$

As anticipated, the magnetic field changes the atom's energy by an amount that depends on the quantum number m . This is why m is often called the **magnetic quantum number**, and it explains the traditional choice of the letter m .

Comparing (9.16) with (9.14) (and remembering that the quantum number m is dimensionless), we see that the quantity in parentheses, $(e\hbar/2m_e)$, must have the dimensions of a magnetic moment. (You should check this directly; see Problem 9.14.) In atomic physics this quantity is a convenient unit for magnetic moments and is called the **Bohr magneton** μ_B , with the value

$$\mu_B = \frac{e\hbar}{2m_e} = 9.27 \times 10^{-24} \text{ A} \cdot \text{m}^2 \quad (9.17)$$

In terms of μ_B , we can rewrite (9.16) in the compact form

$$\Delta E = m\mu_B B \quad (9.18)$$

Since m can have the $2l + 1$ values $l, l - 1, \dots, -l$, we see that the $2l + 1$ states of the original degenerate level now have energies that are equally spaced, an energy $\mu_B B$ apart:

$$(\text{separation of adjacent levels}) = \mu_B B \quad (9.19)$$

The result (9.19) shows that the dimensions of μ_B can also be expressed as energy/(magnetic field); that is, the unit $\text{A} \cdot \text{m}^2$ in (9.17) can be replaced by joules/tesla. If we convert the joules to eV, we get the useful result

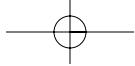
$$\mu_B = 5.79 \times 10^{-5} \text{ eV/T} \quad (9.20)$$

According to (9.19), this means that a field of 1 tesla leads to a separation of adjacent levels by 5.79×10^{-5} eV — a very small separation on the scale of normal atomic levels, which are typically a few eV apart.

Example 9.1

A helium atom is in one of its singlet states (with the two spins antiparallel and hence no spin magnetic moment). One of its electrons is in an s state ($l = 0$) and the other a d state ($l = 2$). The atom is placed in a magnetic field, $B = 2$ T (by normal laboratory standards a fairly strong field). By how much does the magnetic field change the atom's energy?

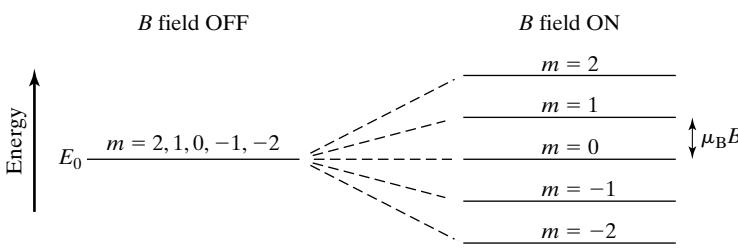
The shift in energy is given by (9.18) as $\Delta E = m\mu_B B$, where, since $l = 2$, the quantum number m can have any of the five values $m = 2, 1, 0,$



294 Chapter 9 • Electron Spin

FIGURE 9.3

The Zeeman effect. In the absence of a magnetic field, an atomic level with $l = 2$ (and no spin magnetic moment) is fivefold degenerate with energy E_0 . When B is switched on, the level splits into a multiplet of five equally spaced levels with separation $\mu_B B$.



$-1, -2$. If the atom is in the state with $m = 0$, its energy is unaltered. If it is in any of the other four states, its energy is shifted as shown in Fig. 9.3. The five resulting energy levels are said to form a multiplet and are evenly spaced above and below the original E_0 , with separation $\mu_B B$. With $B = 2$ tesla, the separation of adjacent levels is

$$\mu_B B = \left(5.79 \times 10^{-5} \frac{\text{eV}}{\text{T}} \right) \times (2 \text{ T}) = 1.2 \times 10^{-4} \text{ eV}$$

We see from this example that even a relatively strong field of a few teslas produces a very small separation of energy levels. Nevertheless, this small splitting of each level into several levels results in an observable splitting of the spectral lines of the light emitted and absorbed by the atom. To illustrate this effect — the Zeeman effect — we consider again the helium atom. Specifically, we consider the ground state and one of the low lying excited states, with energy 21.0 eV above the ground state.

In both of these states of the helium atom, the two electron spins are antiparallel and the resultant spin magnetic moment is zero. Thus the shift in energy produced by a magnetic field is correctly given by (9.18) as

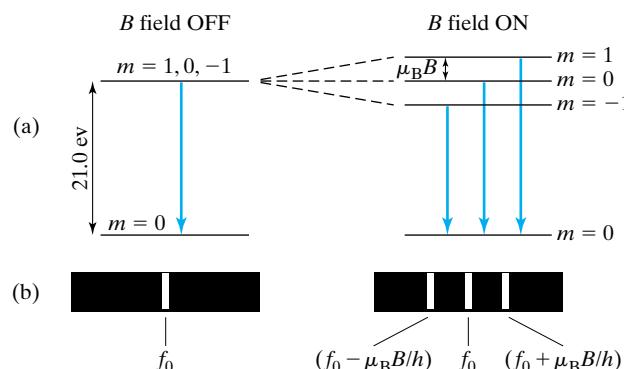
$$\Delta E = m\mu_B B$$

In the ground state both electrons have $l = 0$. Thus, the only possible value of m is $m = 0$, and the shift ΔE is zero. That is, the energy of the ground state is unchanged by a magnetic field. In the excited state, one electron has zero orbital angular momentum while the other has $l = 1$. With $l = 1$, the possible values of m are $m = 1, 0$, and -1 , and the magnetic field splits this level into three equally spaced levels, a distance $\mu_B B$ apart, as shown in Fig. 9.4.

Let us now consider transitions in which a helium atom drops from the excited state just described to the ground state and emits a photon. In the

FIGURE 9.4

(a) The ground state ($l = 0$) and one of the excited levels ($l = 1$) of helium. When a magnetic field is applied, the upper level splits into three, while the ground state is unaffected. (The splitting of the levels is greatly exaggerated since, even in the strongest magnetic fields obtainable in a lab — about 40 T — the separation is only $\mu_B B \approx 2 \times 10^{-3}$ eV.) (b) With the magnetic field on, there are three distinct transitions possible and hence three distinct spectral lines, as shown on the right.



absence of a magnetic field, both levels have unique energies separated by 21.0 eV, and the photon has energy $E_\gamma = 21.0 \text{ eV}$ and frequency $f_0 = E_\gamma/h$. Thus a spectrometer would reveal a single spectral line with frequency f_0 , as indicated at the bottom left of Fig. 9.4. If we now apply a magnetic field B , the upper level splits into three closely spaced levels. Therefore, there are now three possible transitions with three slightly different energies, as indicated by the three downward arrows on the right side of Fig. 9.4. In a gas of excited helium there would normally be atoms in all three levels. Therefore, all three transitions would occur, and a spectrometer would now reveal a triplet of three closely spaced spectral lines with frequencies $f_0 + \mu_B B/h$, f_0 and $f_0 - \mu_B B/h$, as shown at the bottom right of Fig. 9.4.

The Zeeman effect was discovered in 1896, well before the development of quantum mechanics. Naturally, attempts were made to explain it on the basis of classical mechanics. As it happens, the classical theory of the Zeeman effect gives the correct answers for any two atomic states whose spin magnetic moments are zero. Thus a Zeeman splitting like that shown in Fig. 9.4(b) agreed with classical predictions and came to be called the **normal Zeeman effect**.^{*} Unfortunately (for classical mechanics), the effect of a magnetic field on many atoms was found to be much more complicated than this and did not agree with the classical theory. As we will see, these more complicated shifts, which were called **anomalous Zeeman splittings**, involve the spin magnetic moment of the electron and were among the first indications that the electron has a spin.

Example 9.2

What is the wavelength λ_0 of the transition shown on the left of Fig. 9.4? If a magnetic field of 2 T is applied to the helium atom, what are the shifts $\Delta\lambda$ of the outer two spectral lines on the right of Fig. 9.4(b)?

With $B = 0$ the two states are 21.0 eV apart in energy and the emitted photon has $E_\gamma = 21.0 \text{ eV}$. The corresponding wavelength is

$$\lambda_0 = \frac{hc}{E_\gamma} = \frac{1240 \text{ eV} \cdot \text{nm}}{21.0 \text{ eV}} = 59.0 \text{ nm}$$

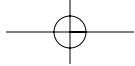
which is in the far ultraviolet.

If we switch on a magnetic field, the upper level splits into the three equally spaced levels shown in Fig. 9.4, separated by energy $\mu_B B$:

$$\begin{aligned} (\text{separation of adjacent levels}) &= \mu_B B \\ &= \left(5.79 \times 10^{-5} \frac{\text{eV}}{\text{T}} \right) \times (2 \text{ T}) \\ &= 1.2 \times 10^{-4} \text{ eV} \end{aligned}$$

The $m = 0$ state has the same energy as when $B = 0$, and the wavelength of photons emitted from this state is unchanged. The energy of photons emitted

^{*}The example of Fig. 9.4(b) involved a transition from $l = 1$ to $l = 0$ and gave a splitting into three spectral lines. One might imagine that higher l values would lead to more than three lines; in fact, however, the normal Zeeman effect always produces exactly three lines (Problem 9.17).



296 Chapter 9 • Electron Spin

from the states with $m = \pm 1$ is changed by $\Delta E_\gamma = \pm \mu_B B$. Since this is small, the shift in their wavelength is well approximated as

$$\begin{aligned}\Delta\lambda &\approx \frac{d\lambda}{dE_\gamma} \Delta E_\gamma = -\frac{hc}{E_\gamma^2} \Delta E_\gamma \\ &= -\frac{1240 \text{ eV} \cdot \text{nm}}{(21.0 \text{ eV})^2} \times (\pm 1.2 \times 10^{-4} \text{ eV}) \\ &= \mp 3.4 \times 10^{-4} \text{ nm}\end{aligned}\quad (9.21)$$

The Zeeman shift of wavelength is so small that the earliest observations could not distinguish the separate spectral lines. At first, all that was detected was a broadening of the original single line; but later experiments with better resolution showed that the line was indeed split into several separate lines. Today, spectrometers can resolve splittings of order 10^{-8} nm, and the Zeeman shifts can be measured very accurately. An important modern application is to measure the splitting of an identified spectral line and hence to find an unknown magnetic field. This is especially useful in astronomy since the magnetic fields of the sun and stars cannot be measured directly.

9.5 Spin Magnetic Moments

We have seen that as an atomic electron orbits around the nucleus, it produces a magnetic moment given by (9.13) as

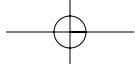
$$\boldsymbol{\mu}_{\text{orb}} = -\frac{e}{2m_e} \mathbf{L} \quad (9.22)$$

If we visualize the electron as a tiny, rigid ball of charge spinning on its axis, we would expect this spinning motion to produce an additional spin magnetic moment. Each piece of the electron would be carried in a circular path around the axis and hence constitute a small current loop. Each such loop would produce a magnetic moment, and the sum of all these moments would be the total spin magnetic moment $\boldsymbol{\mu}_{\text{spin}}$. Since $\boldsymbol{\mu}_{\text{spin}}$ would be proportional to the angular velocity $\boldsymbol{\omega}_{\text{spin}}$, which in turn is proportional to the spin angular momentum \mathbf{S} , we would expect to find $\boldsymbol{\mu}_{\text{spin}} \propto \mathbf{S}$ or

$$\boldsymbol{\mu}_{\text{spin}} = -\gamma \mathbf{S} \quad (9.23)$$

where γ is a constant called the **spin gyromagnetic ratio**. [We have put a minus sign in (9.23) because the electron's charge is negative, and $\boldsymbol{\mu}_{\text{spin}}$ and \mathbf{S} are in opposite directions.] In the case of the orbital motion, the gyromagnetic ratio is seen from (9.22) to be $e/2m_e$. The spin gyromagnetic ratio would not necessarily have this same value, since it would depend on the distributions of charge and mass within the electron. If the charge were concentrated farther out than the mass, then $\mu_{\text{spin}}/\mathbf{S}$ would be relatively large; if the charge were concentrated nearer the center, $\mu_{\text{spin}}/\mathbf{S}$ would be smaller.

The classical picture of the electron as a rigid spinning ball of charge is not strictly correct. For example, if the radius of the ball is taken consistent with modern observations, the equatorial speed turns out to be greater than c , which is impossible (Problem 9.8). Nevertheless, the conclusions that there

Section 9.6 • The Anomalous Zeeman Effect **297**

should be a magnetic moment with the general form (9.23) and that the spin gyromagnetic ratio does not necessarily have the same value as the orbital ratio $e/2m_e$ are both correct. Experiment shows that there is a magnetic moment with the form (9.23) and that the spin gyromagnetic ratio γ is e/m_e , just twice* the value of the orbital ratio, $e/2m_e$; that is,

$$\boldsymbol{\mu}_{\text{spin}} = -\frac{e}{m_e} \mathbf{S} \quad (9.24)$$

The total magnetic moment of any electron is just the sum of its orbital and spin moments

$$\boldsymbol{\mu}_{\text{tot}} = \boldsymbol{\mu}_{\text{orb}} + \boldsymbol{\mu}_{\text{spin}} = -\frac{e}{2m_e} (\mathbf{L} + 2\mathbf{S}) \quad (9.25)$$

As we describe in the next two sections, much of the evidence for the electron's spin comes from the repeated success of the formula (9.25) in explaining a wide variety of experimental results.

The suggestion that the electron has a spin angular momentum and a corresponding magnetic moment [given, as we now know, by (9.24)] is generally credited to the Dutch physicists Samuel Goudsmit and George Uhlenbeck (1925). Their suggestion was based on an analysis of the anomalous Zeeman effect (which we describe in Section 9.6) and of the fine structure in atomic spectra (Section 9.7). However, it is worth mentioning that similar suggestions had been made by other physicists. In particular, Arthur Compton had suggested that a spin magnetic moment for the electron could possibly explain the phenomenon of ferromagnetism, a suggestion that later proved to be correct.

**George Uhlenbeck
and Samuel
Goudsmit**

(1900–1988, Dutch, American)
(1902–1978, Dutch, American)



Uhlenbeck (center) and Goudsmit (right) are shown here with colleague Oskar Klein (left). In 1925 Goudsmit and Uhlenbeck, while both graduate students at Leiden, showed that several puzzles in atomic spectra could be explained if the electron was assumed to have a spin angular momentum with quantum number $s = \frac{1}{2}$. Both moved to the United States in 1927, and both worked at Michigan and then MIT. Goudsmit became editor of the *Physical Review*.

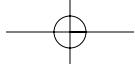
9.6 The Anomalous Zeeman Effect

When an atom is placed in a magnetic field, its energy levels undergo small shifts and individual levels get split into several closely spaced levels. This results in a splitting of the spectral lines into closely spaced “multiplets” of lines — an effect known as the Zeeman effect.

In Section 9.4 we calculated in detail the Zeeman splitting for the so-called singlet states of the helium atom (the states in which the two spin magnetic moments cancel out and can therefore be ignored). The results of those calculations are correct for any atomic state in which the spin magnetic moments cancel. In general, however, the Zeeman effect does *not* agree with the splittings calculated in Section 9.4, but does agree with a corresponding calculation using the correct magnetic moment (9.25), including both orbital and spin moments. For historical reasons, the splitting of levels and spectral lines is called the *normal Zeeman effect* in those cases where spin has no effect and called the *anomalous Zeeman effect* in those cases where spin does contribute.

The correct calculation of the anomalous Zeeman splitting is quite complicated, depending as it does on both orbital and spin moments. To simplify our discussion, we consider here the simple case of a hydrogen atom in a state

*Precise measurements show that the spin ratio is actually not exactly twice the orbital value (instead of 2, the factor is 2.0023). The observed value is successfully predicted by the relativistic quantum theory called quantum electrodynamics.



298 Chapter 9 • Electron Spin

with no orbital angular momentum; that is, an s state, with $l = 0$. If the electron had no spin, then, with $l = 0$, the atom would have no magnetic moment at all, and would be completely unaffected by a magnetic field. In fact, of course, the electron does have spin and, even though $l = 0$, there is a magnetic moment

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{\text{spin}} = -\frac{e}{m_e} \mathbf{S} \quad (9.26)$$

When a magnetic field \mathbf{B} (in the z direction) is switched on, the energy changes by an amount

$$\Delta E = -\boldsymbol{\mu} \cdot \mathbf{B} = \frac{e}{m_e} S_z B$$

Since the possible values of S_z are

$$S_z = \pm \frac{1}{2}\hbar$$

it follows that

$$\Delta E = \pm \frac{e\hbar}{2m_e} B = \pm \mu_B B \quad (9.27)$$

If the electron has spin up, its energy is raised by $\mu_B B$; if it has spin down, its energy is lowered by the same amount. The resulting separation of the two levels is therefore

$$(\text{separation of levels}) = 2\mu_B B \quad (9.28)$$

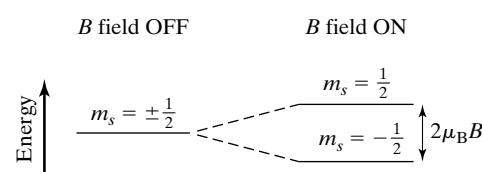
Notice that this separation is twice the value (9.19) predicted for the normal Zeeman effect; this is because the spin gyromagnetic ratio is twice the orbital ratio.

We conclude that any $l = 0$ level in hydrogen should be split into two neighboring levels by a magnetic field. This splitting is sketched in Fig. 9.5. That these levels are observed to split into two levels is strong evidence that the electron has a spin $s = \frac{1}{2}$ (and hence just two possible orientations). The agreement of the observed level separation with (9.28) confirms the expression (9.26) for $\boldsymbol{\mu}_{\text{spin}}$.

The Zeeman effect has been observed in many different levels of dozens of different atoms, and in all cases the results confirm that the electron's total magnetic moment is given by (9.25) as $-(e/2m_e)(\mathbf{L} + 2\mathbf{S})$, that the angular momentum vector \mathbf{S} has magnitude $\sqrt{s(s+1)}\hbar$ with s always equal to $\frac{1}{2}$, and that S_z has the two possible values $\pm \frac{1}{2}\hbar$.

FIGURE 9.5

Any s level in hydrogen is split by a B field into two levels because of the electron's spin magnetic moment. The separation of the levels is $2\mu_B B$.



9.7 Fine Structure*

*Though important, this material will not be needed later and can be omitted without loss of continuity.

The Zeeman effect is a splitting of atomic energy levels caused by an *externally applied* magnetic field. In most atoms there is a permanent *internal* magnetic field due to the motion of the charges inside the atom. Even when there is no external magnetic field, this internal field can cause a small splitting of the energy levels and, hence, of the atomic spectrum. These splittings due to the internal magnetic field are called **fine structure**. As an illustration, we describe briefly the fine structure of hydrogen, many of whose spectral lines were found to be *doublets* consisting of two closely spaced lines.

Let us consider the states of a hydrogen atom with some definite energy (given by quantum number n) and definite nonzero orbital angular momentum (given by quantum number l). We can understand the fine structure of these states from the following semiclassical argument: In the rest frame of the electron, the proton orbits around the electron, as shown in Fig. 9.6. Therefore, the electron finds itself in the magnetic field produced by the current loop of an orbiting positive charge. This field is proportional to the orbital frequency of the proton (as seen in the electron's rest frame), which in turn is proportional to the orbital angular momentum \mathbf{L} of the electron, as seen in the proton's rest frame. Therefore, the electron sees a field \mathbf{B} that is proportional to \mathbf{L} :

$$\mathbf{B} \propto \mathbf{L} \quad (9.29)$$

As can be seen in Fig. 9.6, the direction of \mathbf{B} is the same as that of \mathbf{L} . Therefore, the constant of proportionality in (9.29) is positive.

Since the electron has a magnetic moment* $\boldsymbol{\mu}_{\text{spin}} = -(e/m_e)\mathbf{S}$, the magnetic field \mathbf{B} gives it an additional energy

$$\Delta E = -\boldsymbol{\mu}_{\text{spin}} \cdot \mathbf{B} = \frac{e}{m_e} \mathbf{S} \cdot \mathbf{B} \propto \mathbf{S} \cdot \mathbf{L} \quad (9.30)$$

Because it is proportional to $\mathbf{S} \cdot \mathbf{L}$, this energy is often described as the **spin-orbit** energy. We will see that the spin-orbit energy is usually very small by atomic standards. (Typically, it is a very small fraction of an eV.) This means that the hydrogen energy levels calculated in Chapter 8, which ignored all effects of spin, are still an excellent approximation. Nevertheless, the correction (9.30) does cause a small splitting of the levels, as we now show.

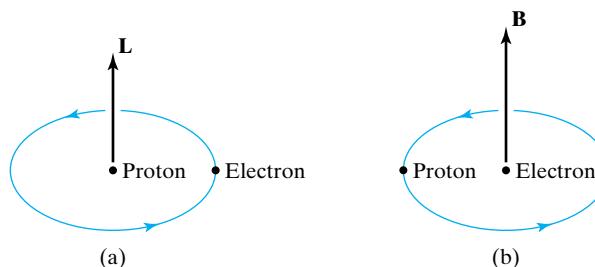
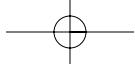


FIGURE 9.6

- (a) In the proton's rest frame the electron orbits around the proton, with orbital angular momentum \mathbf{L} . (Because the proton is so much heavier than the electron, this frame is very close to the rest frame of the atom as a whole, and is the frame usually considered.)
- (b) In the electron's rest frame, the proton orbits around the electron. The sense of the orbit is the same in both pictures. Since the proton's charge is positive, it produces a magnetic field \mathbf{B} (given by the right-hand rule) in the direction shown. Therefore, \mathbf{B} is in the same direction as \mathbf{L} .

*We need consider only the spin magnetic moment because the orbital momentum is always zero in the electron's rest frame.



300 Chapter 9 • Electron Spin

According to (9.30), the electron has a magnetic energy that depends on the orientation of \mathbf{S} relative to \mathbf{L} . The spin \mathbf{S} can have two possible orientations relative to any definite direction. Therefore, the term $\mathbf{S} \cdot \mathbf{L}$ in (9.30) can have two possible different values. This means that the states with any particular values of n and l actually belong to two slightly different energy levels. Those states in which \mathbf{S} is parallel to \mathbf{L} have a slightly higher energy; those in which \mathbf{S} is antiparallel to \mathbf{L} are slightly lower.

The argument just given applies to any state with $l \neq 0$. If $l = 0$, then since $\mathbf{B} \propto \mathbf{L}$, it follows that the B field seen by the electron is zero and hence that there is no spin-orbit splitting for s states.

The splitting of each level (with $l \neq 0$) into two levels implies a corresponding splitting of the spectral lines of hydrogen. As an example, let us consider transitions in which a hydrogen atom in one of its $2p$ states drops to the ground state. We have seen that the ground state is not split by the spin-orbit interaction, whereas the $2p$ states are split into two levels. Thus the energy-level diagram for these states is as shown in Fig. 9.7, and there are two different possible photon energies. Therefore, the transitions from $2p$ states to the ground state produce a doublet of spectral lines, as shown in Fig. 9.7(b).

To estimate the separation of these two lines, we must first find the magnetic field B of the orbiting proton, as seen by the $2p$ electron. A straightforward calculation (Problem 9.21) shows that

$$B \approx 0.39 \text{ T} \quad (9.31)$$

According to (9.28), this implies that the $2p$ levels are separated by an energy,

$$\begin{aligned} (\text{separation of } 2p \text{ levels}) &= 2\mu_B B \\ &\approx 2 \times \left(5.8 \times 10^{-5} \frac{\text{eV}}{\text{T}} \right) \times (0.39 \text{ T}) \\ &\approx 4.5 \times 10^{-5} \text{ eV} \end{aligned} \quad (9.32)$$

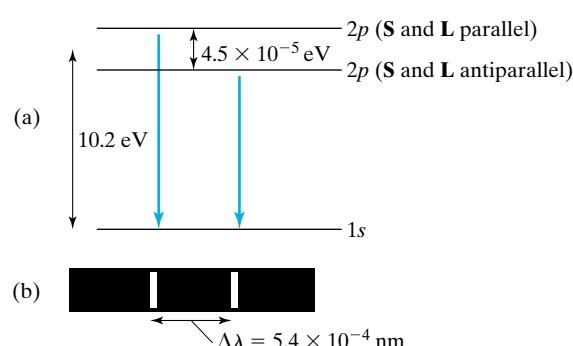
This is extremely small compared to the distance between the $2p$ and $1s$ levels, which is $(13.6 - 3.4) = 10.2 \text{ eV}$. Therefore, the difference in wavelengths of the emitted photons is approximately

$$\begin{aligned} \Delta\lambda &\approx \left| \frac{d\lambda}{dE_\gamma} \right| \Delta E_\gamma = \frac{hc}{E_\gamma^2} \Delta E_\gamma \\ &= \frac{1240 \text{ eV} \cdot \text{nm}}{(10.2 \text{ eV})^2} \times (4.5 \times 10^{-5} \text{ eV}) \\ &= 5.4 \times 10^{-4} \text{ nm} \end{aligned}$$

as indicated in Fig. 9.7.

FIGURE 9.7

Fine structure in hydrogen. (a) The $1s$ states have a unique energy, while the $2p$ states belong to two slightly different energy levels, those with \mathbf{S} parallel to \mathbf{L} being slightly higher. (The separation of these levels is exaggerated by a factor of 50,000.) (b) Therefore, transitions from $2p$ to $1s$ involve photons with two slightly different energies and produce a doublet of spectral lines, as shown.



It should be emphasized that in our discussion of fine structure we have treated the electron as a classical orbiting particle. Also, we have consistently treated the hydrogen atom nonrelativistically, and although this is certainly an excellent approximation, there are small corrections required to allow for relativity. It turns out that these relativistic corrections are of the same order of magnitude as the spin-orbit energy discussed here. (See Problems 9.21 to 9.23.) Thus a correct analysis of the fine structure of hydrogen needs to be fully quantum mechanical and to take account of relativity. Under these conditions our calculation of the splittings can only be regarded as an order-of-magnitude estimate. That the answer (9.32) is correct to two significant figures is simply a happy accident. Nevertheless, all of our general conclusions are qualitatively correct.

9.8 Magnetic Resonance Imaging (MRI)*

*Though important, this material will not be needed later and can be omitted without loss of continuity.

In the 1970s, scientists from several universities and industries developed a new medical diagnostic technique called **magnetic resonance imaging** (MRI) or nuclear magnetic resonance (NMR). Today, it is a common and safe imaging technique, producing pictures of internal organs without exposing the patient to ionizing radiation, such as conventional X-rays. (See Fig. 9.8) Although more expensive and slower than X-rays, MRI is better able to distinguish between different types of soft tissues. This remarkable technique involves measuring the concentration of *proton* spins in the patient's body.

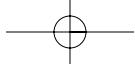
We begin with a short discussion of proton spin. As we mentioned at the end of Section 9.2, the proton, like the electron, is a spin-half particle. Like the electron, the proton has S_z equal to $\pm \frac{1}{2}\hbar$. Also like the electron, it has a magnetic moment related to its spin by an equation similar to (9.23),

$$\boldsymbol{\mu}_{\text{proton}} = \gamma_{\text{proton}} \mathbf{S} \quad (9.33)$$



FIGURE 9.8

A magnetic resonance imaging (MRI) apparatus. The large structure contains a superconducting magnet producing a 1-tesla magnetic field. The smaller cylindrical structure surrounding the head of the patient is the pick-up coils, which detects the radio-frequency signal from the protons in the patient's body.



302 Chapter 9 • Electron Spin

There are two differences between this equation and (9.23) for electrons: (1) In this equation there is no minus sign because the proton is positive. (2) The gyromagnetic factor for protons is about 2600 times smaller than for electrons, resulting in a proton magnetic moment smaller by 2600 than the electron moment. The smallness of the nuclear gyromagnetic factor is due to the largeness of the proton mass: Just as the electron gyromagnetic ratio is given by $\gamma = e/m_e$, the proton gyromagnetic ratio is given approximately* by $\gamma = e/m_p$, where m_p is the proton mass.

Just as with the electron, a proton placed in a magnetic field can exist in one of two different energy states, corresponding to the magnetic moment and B field being aligned or anti-aligned. Looking at (9.33) (and dropping the “proton” subscripts), we see that the energies of the two levels are given by

$$-\mu_z B = \gamma S_z B = \pm \frac{1}{2} \gamma \hbar B \quad (9.34)$$

and the energy-level separation is $\Delta E = \gamma \hbar B$. Transitions between these two levels can be induced by the absorption or emission of electromagnetic radiation of angular frequency ω such that $\hbar \omega = \Delta E = \hbar \gamma B$. When the magnetic field is a few tesla, this frequency happens to be in the range of radio frequencies, which is roughly 1 to 100 MHz. Specifically, for a 1-tesla field, the frequency of this radiation is $f = \omega/2\pi = 42.6$ MHz. (See Problem 9.25)

Humans, of course, are full of protons. Every water molecule (H_2O) in our body contains two hydrogen nuclei, which are protons. (We will see that the other nuclei, such as oxygen, are not normally involved in MRI.) The human patient, with all her proton spins, is placed in a very strong magnetic field, typically $B = 1$ or 2 tesla. Because of this B field, every proton in the patient’s body exists in one of two energy levels, with spin up or spin down. In an MRI apparatus, a transmitter coil surrounding the patient produces a pulse of radio waves at the correct frequency, the “resonant frequency,” to induce transitions between the spin levels, generating a nonequilibrium population of spin up versus spin down in the patient. As the spins spontaneously relax back to equilibrium, the transitions between the two levels produce electromagnetic fields with a specific radiofrequency precisely proportional to the applied field, $\omega = \gamma B$. This electromagnetic signal is detected by a receiver coil around the patient,[†] and the frequency of this signal gives a very accurate measurement of the magnetic field in which the patient has been placed.

So far we have described a very expensive way to use the protons in a human to measure a magnetic field. Now comes the clever part. The magnetic field in which the patient lies is not uniform; it is intentionally made inhomogeneous so that the field varies linearly from one side of the patient to the other. The protons in the high-field side of the patient have greater energy-level splitting and produce higher-frequency radio emissions than protons of the low-field side. Thus, a particular proton’s radio signal contains position information; the signal’s frequency is exactly proportional to the proton’s position in one dimension, the axis along which the field varies. By switching the axis of the field inhomogeneity among the x , y , and z directions while monitoring the radio signal, one can collect enough information to construct a three-dimensional image of

*More precisely, the magnetic moment of a proton is $1.40e/m_p$. The calculation of the factor of 1.40 is quite complicated and arises from the complex internal structure of the proton, which consists of 3 quarks.

[†] Often, a single coil acts as both transmitter and receiver. This transceiver coil is separate from the superconducting coils of the magnet producing the large B field.

**FIGURE 9.9**

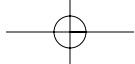
(a) MRI scan of the head and (b) of the spine.

the proton density in the patient. The details are quite messy: The receiver coil around the patient is picking up a multitude of different frequencies from all the protons at once. Disentangling the frequencies to produce an image requires very sophisticated software and significant computer power.* But the results are remarkable; images with a resolution 1 to 0.1 mm can be acquired in several minutes. Two examples are shown in Fig. 9.9. Some further details of the MRI process are presented in Problems 9.26 and 9.27.

Of the many other types of nuclei in the patient's body, such as O, N, C, and so forth, each has its own value of spin angular momentum, and each has a characteristic gyromagnetic ratio and magnetic moment. In some cases, such as ^{16}O , the spin happens to be zero, but in many cases the magnetic moment is nonzero and these nuclei can produce an MRI signal. However, all these other nuclei possess gyromagnetic ratios γ and corresponding frequencies $\omega = \gamma B$ that are different from that of the proton. The transmitter/receiver electronics in an MRI apparatus is tuned to induce transitions only for nuclei with frequencies near the proton frequency. Consequently, the presence of the other nuclei does not affect the proton signal. Of the many types of nuclei in humans, hydrogen nuclei are chosen for MRI imaging because they are both numerous and have a large magnetic moment, so they provide the strongest signal.

The single most expensive component of an MRI apparatus is the magnet. Producing such a well-controlled field of 1 or 2 tesla over a volume large enough to insert a human is a major engineering accomplishment. The magnet used in an MRI apparatus has superconducting windings cooled to a few degrees above absolute zero with the use of liquid helium. (See Chapter 14 for a discussion of superconductivity.) Although such superconducting magnets are quite expensive (roughly a million dollars), they are cheaper to operate than conventional iron-core electromagnets, which would require enormous

*When MRI units were first used in the late 1970s, large mainframe computers were needed to handle the image processing calculations. Today (2003), one good PC can handle the job.



304 Chapter 9 • Electron Spin

amounts of electrical power and cooling water to produce a magnetic field of comparable strength. The commercial production of such reliable, affordable superconducting magnets was achieved in the late 1970s and was a major milestone on the road to making MRI a practical tool.

CHECKLIST FOR CHAPTER 9

CONCEPT	DETAILS
Spin angular momentum of the electron	quantum number $s = \frac{1}{2}$ magnitude $S = \sqrt{s(s+1)}\hbar = \sqrt{3}\hbar/2$ (9.2) $S_z = \pm \frac{1}{2}\hbar$ (9.5)
Classical magnetic moment μ potential energy	$\mu = i\mathbf{A}$ (9.8) $U = -\mu \cdot \mathbf{B}$ (9.10)
Gyromagnetic ratio γ	ratio of magnetic moment μ to angular momentum (Secs. 9.3 and 9.5)
Electron's orbital magnetic moment	$\mu_{\text{orb}} = -(e/2m_e)\mathbf{L}$ (9.13)
The Bohr magneton	convenient unit for atomic magnetic moments, $\mu_B = e\hbar/(2m_e)$ (9.17)
Zeeman splitting	$\Delta E = m\mu_B B$ (9.18), energy shift of orbital magnetic moment in a field
Electron's spin magnetic moment	$\mu_{\text{spin}} = -(e/m_e)\mathbf{S}$ (9.24)
Fine structure *	$\Delta E = -\mu_{\text{spin}} \cdot \mathbf{B} \propto \mathbf{S} \cdot \mathbf{L}$, where \mathbf{B} is due to orbital motion of proton as seen by electron (9.30)
Magnetic resonance imaging (MRI or NMR) *	proton spin in a magnetic field, radiofrequency emissions of frequency $\omega = \gamma B$ (Sec. 9.8)

PROBLEMS FOR CHAPTER 9

SECTION 9.2 (Spin Angular Momentum)

- 9.1** • One can visualize the quantized values of the spin angular momentum \mathbf{S} with a semiclassical “vector model,” as described in Section 8.6 for the orbital angular momentum \mathbf{L} . In particular, the quantization of S_z requires that the vector \mathbf{S} must lie on certain cones, like the one sketched in Fig. 8.15. **(a)** Make a sketch similar to Fig. 8.14 showing the two possible orientations of \mathbf{S} for an electron. **(b)** What is the angle between \mathbf{S} and the z axis for these two states?
- 9.2** • There exist subatomic particles with spin magnitudes different from that of the electron. However, in all cases they obey the same rules: The magnitude of \mathbf{S} is $\sqrt{s(s+1)}\hbar$, where s is a fixed number, integer or half-integer; and the possible values of S_z are $m_s\hbar$, where m_s has the values $s, s-1, \dots, -s$. **(a)** For a particle with $s = \frac{3}{2}$, how many different values of S_z are there, and what are they? **(b)** Draw a vector model diagram similar to Fig. 8.14 showing the possible orientations of \mathbf{S} . **(c)** What is the minimum possible angle between \mathbf{S} and the z axis?
- 9.3** • Answer the same questions as in Problem 9.2, but for a particle with spin quantum number $s = 1$.

9.4 • Make a table showing the values of the four quantum numbers n, l, m, m_s for each of the 18 states of the hydrogen atom with energy $E = -E_R/9$.

9.5 • Make a table showing the values of the four quantum numbers n, l, m, m_s and the energies for each of the 10 lowest-lying quantum states (not energy levels) of a hydrogen atom.

9.6 • Compute the ratio S/L of the magnitudes of the spin angular momentum to the orbital angular momentum for **(a)** an s electron, **(b)** a p electron, **(c)** a d electron, **(d)** an f electron. Note that for all but the s electron, the size of the spin angular momentum is of the same order of magnitude as the orbital angular momentum.

9.7 •• **(a)** Write an expression for the magnitude of orbital angular momentum L_{orb} of the earth due to its orbital motion about the sun. **(b)** Assume that the earth has uniform mass density, and derive an expression for the spin angular momentum L_{spin} of the earth due to its rotation about its axis. **(c)** Show that for the earth, the ratio $L_{\text{spin}}/L_{\text{orb}}$ is given by the expression $2R_E^2 T_{\text{year}}/(5R_{SE}^2 T_{\text{day}})$, where R_E is the earth’s radius, R_{SE} is the earth-sun distance, T_{year} is one year, and

T_{day} is one day. Compute the value of this expression, and note that $L_{\text{spin}}/L_{\text{orb}} \ll 1$.

- 9.8** •• We have said that a classical picture of the electron as a spinning ball of matter is unsatisfactory. To illustrate this, consider the following: Modern measurements show that the electron's radius is certainly less than 10^{-18} m. Write an expression for the angular momentum of a uniform spinning ball of mass m_e , radius r , and equatorial speed v . By equating this to the observed spin $\sqrt{3}\hbar/2$, find the minimum possible value of v . What is v/c ?

SECTION 9.3 (Magnetic Moments)

- 9.9** • A current of 0.4 A flows around a single circular loop of radius 1 cm. (a) What is the resulting magnetic moment, μ ? (b) If the loop is placed in a magnetic field $B = 1.5$ T, with μ perpendicular to \mathbf{B} , what is the torque on the loop? (c) What is the difference in energy between the cases that μ is parallel to \mathbf{B} and antiparallel?
- 9.10** • The SI units of magnetic moment μ are ampere · meter². According to (9.13), the magnetic moment of an orbiting electron is

$$\mu = -\frac{e}{2m_e} \mathbf{L} \quad (9.35)$$

Verify that this has the correct units.

- 9.11** • A typical atomic magnetic moment is of order 10^{-23} A · m². Assuming that this is the result of a current i circulating around a single circular loop of radius 0.1 nm (a typical atomic radius), how big is i ?
- 9.12** •• The energy of a magnetic moment μ in a magnetic field B pointing along the z axis is $-\mu_z B$. For an electron in orbit around a proton, μ_z is given by (9.35) as $-(e/2m_e)L_z$. If $B = 10$ T (a large field by the standards of most laboratories) and if the electron is in a p state with $L_z = \hbar$, what is the magnetic energy due to the orbital magnetic moment (in joules and in eV)?

SECTION 9.4 (The Zeeman Effect)

- 9.13** • (a) Using the known SI values of e , \hbar , and m_e , find the SI value of the Bohr magneton $\mu_B = e\hbar/2m_e$. (b) Given that the units ampere · meter² are the same thing as joule/tesla, find μ_B in eV/tesla.
- 9.14** •• (a) Verify that the Bohr magneton $\mu_B = e\hbar/2m_e$ has the units of magnetic moment: namely, ampere · meter². (b) Verify that the units ampere · meter² are the same thing as joule/tesla. (Remember that the tesla is the unit of B field, defined by the Lorentz-force equation $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$.)
- 9.15** •• A helium atom is in an energy level with one electron occupying an s state ($l = 0$) and the other an f state ($l = 3$). The two electron spins are antiparallel so that the spin magnetic moments cancel. The atom is placed in a magnetic field $B = 0.8$ T. (a) Sketch the resulting splitting of the original energy level. (b) What is the energy difference between adjacent levels of the resulting multiplet?

- 9.16** •• Imagine a hydrogen atom in which the electron has no spin [so that the only magnetic moment is the orbital magnetic moment given by (9.35)]. The atom is placed in a magnetic field $B = 1.5$ T along the z axis. (a) Describe the effect of the B field on the $1s$ and $2p$ states of the hypothetical atom. Sketch the energy levels. (b) When $B = 0$, there is a single spectral line corresponding to the $2p \rightarrow 1s$ transition. How many lines does this become when B is switched on? (c) What is the fractional separation $\Delta f/f_0$ between adjacent lines?

- 9.17** •• Consider two levels of the helium atom in both of which the spins are antiparallel and one electron is in an s state ($l = 0$). In the higher level the second electron occupies a d state ($l = 2$), and in the lower level it occupies a p state ($l = 1$). (a) Sketch the splitting of both levels resulting from a magnetic field along the z axis. (b) Imagine a transition from one of the d states, with $L_z = m_i\hbar$, to one of the p states, with $L_z = m_f\hbar$. Since m_i can be 2, 1, 0, -1, or -2 and m_f can be 1, 0, or -1, there are 5×3 or 15 distinct conceivable transitions. How many different photon energies would these 15 transitions produce? (c) Not all of these 15 transitions occur. In fact, it is found that the only transitions observed are those for which

$$(m_f - m_i) = 1, \text{ or } 0, \text{ or } -1 \quad (9.36)$$

(A restriction like this on the transitions that take place is called a *selection rule*, as we discuss in Chapter 11.) Prove that because of the restriction (9.36), there are only *three* distinct photon energies produced in all possible transitions. (This means that the normal Zeeman effect always produces just three spectral lines, however large the angular momenta involved.)

SECTION 9.5 (Spin Magnetic Moments)

- 9.18** • The electron's total magnetic moment μ is given by (9.25). (a) What are the possible values of μ_z for an electron with $l = 0$? (b) Compare these with the values of μ_z for a hypothetical spinless electron with $l = 1$.

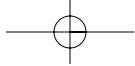
SECTION 9.6 (The Anomalous Zeeman Effect)

- 9.19** • Consider a hydrogen atom in its ground level, placed in a magnetic field of 0.7 T along the z axis. (a) What is the energy difference between the spin-up and spin-down states? (b) An experimenter wants to excite the atom from the lower to the upper state by sending in photons of the appropriate energy. What energy is this? What is the wavelength? What kind of radiation is this? (Visible? UV? etc.)

- 9.20** •• Consider a hydrogen atom in the $3d$ state with $L_z = 2\hbar$ and $S_z = \frac{1}{2}\hbar$. How much does its energy change if it is placed in a magnetic field $B = 0.6$ T along the z axis? [Hint: The total magnetic moment is given by (9.25), and the energy shift is $\Delta E = -\mu_{\text{tot}} \cdot \mathbf{B}$.]

SECTION 9.7 (Fine Structure*)

- 9.21** •• The fine structure of an atomic spectrum results from the magnetic field "seen" by an orbiting electron. In this question you will make a semiclassical estimate of the B field seen by a $2p$ electron in



306 Chapter 9 • Electron Spin

hydrogen. The B field at the center of a circular current loop, i , of radius r is known to be $B = \mu_0 i / 2r$. (a) Treating the electron and proton as classical particles in circular orbits (each as seen by the other), show that the B field seen by the electron is

$$B = \frac{\mu_0}{4\pi} \frac{eL}{m_e r^3} \quad (9.37)$$

where L is the electron's orbital angular momentum ($L = m_e v r$ for a circular orbit). Remember that the current produced by the orbiting proton is $i = ev/2\pi r$, where v is the speed of the proton as seen by the electron (or vice versa). (b) For a rough estimate, you can give L and r their values for the $n = 2$ orbit of the Bohr model, $L = 2\hbar$ and $r = 4a_B$. Show that this gives $B \approx 0.39$ T and hence that the separation, $2\mu_B B$, of the two $2p$ levels is about 4.5×10^{-5} eV.

It should be clear that this semi-classical calculation is only a rough estimate. You have used the Bohr values for L and r . If, for example, you had used the quantum value $L = \sqrt{2}\hbar$, this would have changed your answer by a factor of $\sqrt{2}$. There is another very important reason that the argument used here is only roughly correct: The electron's rest frame is noninertial (since it is accelerated) and a careful analysis by the British physicist L. H. Thomas showed that the energy separation calculated here should include an additional factor of $\frac{1}{2}$. That our answer, 4.5×10^{-5} eV, is correct to two significant figures is just a lucky accident.

- 9.22** •• (a) Use Eq. (9.37) (with the Bohr values $L = 2\hbar$ and $r = 4a_B$) to show that the fine-structure separation $\Delta E_{FS} = 2\mu_B B$ of the two $2p$ levels of hydrogen can be written as

$$\Delta E_{FS} = \frac{m_e (ke^2)^4}{32\hbar^4 c^2} \quad (9.38)$$

[Hint: Since $\mu_0 \epsilon_0 = 1/c^2$ and $k = 1/4\pi\epsilon_0$, you can replace μ_0 by $\mu_0 = 4\pi k/c^2$.]

(b) Show that you can rewrite (9.38) as

$$\Delta E_{FS} = \frac{\alpha^2 E_R}{16} \quad (9.39)$$

where α is the dimensionless *fine-structure constant*

$$\alpha = \frac{ke^2}{\hbar c} \quad (9.40)$$

(c) Show that $\alpha \approx 1/137$, which, together with (9.39), shows that fine structure is indeed a small effect.

- 9.23** •• In Problem 5.10 it was shown that the speed of an electron in a Bohr orbit is of order $v \approx \alpha c$, where α is the fine-structure constant (9.40). Using this value, you can estimate the importance of relativistic corrections to the hydrogen energies, as follows: (a) Write down the correct relativistic expression for the electron's kinetic energy, and use the binomial series (Appendix B) to show that

$$K \approx \frac{1}{2} mv^2 + \frac{3}{8} \frac{mv^4}{c^2}$$

provided that $v \ll c$ (which is certainly true if $v \approx \alpha c$). Thus the relativistic correction ΔE_{rel} to the energy is about $3mv^4/8c^2$. (b) Substitute $v \approx \alpha c$ and show that this gives

$$\Delta E_{rel} \approx \frac{3\alpha^2 E_R}{4}$$

Comparing this with (9.39), we see that relativistic corrections to the hydrogen energy are of the same order as the spin-orbit correction. Therefore, a correct treatment of fine structure must include both effects.

SECTION 9.8 (Magnetic Resonance Imaging *)

- 9.24** • Just as the electron magnetic moment is given approximately by the Bohr magneton (9.17), the nuclear magnetic moment is given approximately by the so-called **nuclear magneton**, defined as

$$\mu_N = \frac{e\hbar}{2m_p} \quad (9.41)$$

where m_p is the proton mass. (a) Compute the value of the nuclear magneton in units of eV/T. Compare this with the size of the Bohr magneton. (b) The magnetic moment of the proton has magnitude $2.7\mu_N$. Compute the size of the energy-level separation in eV of a proton placed in a magnetic field $B = 1$ T. Compare this energy with the size of the thermal energy, $k_B T$, at room temperature.

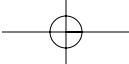
- 9.25** • Use the results of the previous problem to show that for a proton in a $B = 1$ -tesla field, the energy-level splitting corresponds to a frequency of 42.6 MHz.

- 9.26** •• According to statistical mechanics (Ch. 15), if a particle is in thermal equilibrium at temperature T and it can exist in one of two quantum states with energies E_1 and E_2 , then the probabilities $P(E_1)$ and $P(E_2)$ that the particle will occupy each of the two levels are related by

$$\frac{P(E_2)}{P(E_1)} = \frac{e^{-E_2/kT}}{e^{-E_1/kT}}$$

For a proton in a magnetic field $B = 1$ T at temperature $T = 300$ K, compute the ratio $P(\text{higher energy})/P(\text{lower energy})$. Note how close to 1 this result is. Because of thermal agitation, a collection of protons is only very weakly magnetized when placed in a strong magnetic field; that is, there is only a very slight excess of moments aligned with the field over those anti-aligned. Consequently, the proton radio signal in MRI is exceedingly weak, and several minutes of signal averaging are required to produce a high-quality image.

- 9.27** ••• *Semi-classical model of spin in a magnetic field.* A large collection of proton spins in a magnetic field acts, in many ways, like a classical magnetic moment μ . Given that the torque on a moment in a field is given by $\Gamma = \mu \times \mathbf{B}$, and that the classical angular momentum \mathbf{L} is related to the moment μ by $\mu = \gamma \mathbf{L}$, and finally that $\Gamma = d\mathbf{L}/dt$, show that a moment μ in a field \mathbf{B} will *precess* about the field direction with an angular frequency given by $\omega = \gamma B$.



Chapter 10

Multielectron Atoms; the Pauli Principle and Periodic Table

- 10.1** Introduction
- 10.2** The Independent-Particle Approximation
- 10.3** The IPA Energy Levels
- 10.4** The Pauli Exclusion Principle
- 10.5** Fermions and Bosons; the Origin of the Pauli Principle*
- 10.6** Ground States of the First Few Elements
- 10.7** The Remaining Elements
- 10.8** The Periodic Table
- 10.9** Excited States of Atoms*

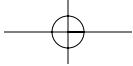
Problems for Chapter 10

* These sections can be omitted without significant loss of continuity.

10.1 Introduction

In Chapter 8 we saw how the Schrödinger equation was used to predict the properties of the hydrogen atom. Another early triumph for the Schrödinger equation was that, unlike the Bohr model, it could be extended successfully to atoms with more than one electron. Today it appears that the Schrödinger equation can correctly account for the structure of all of the hundred or so multielectron atoms and for the way in which these atoms come together to form molecules. Thus, in principle at least, the Schrödinger equation can explain all of chemistry. It is this impressive accomplishment that we introduce in this chapter. We complete the story in Chapter 12 where we describe how atoms combine to form molecules.

Our first step is to find a way to handle quantum systems containing several particles. The simplest method — and the method we describe here — is called the independent-particle approximation. When this method was first applied to multielectron atoms, it predicted many atomic levels, some of which exist but many of which do not. This partial failure led to the discovery of a new law that applies to multielectron systems. We describe this law, called the Pauli exclusion principle, in Section 10.4. Armed with this principle, we can successfully describe the general properties of all multielectron atoms from helium through uranium and beyond, as we sketch in Sections 10.6 to 10.9.

**308** Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

10.2 The Independent-Particle Approximation

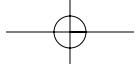
The Schrödinger equation for one electron moving around a nucleus can be solved exactly, as we described in Chapter 8. With two or more electrons, an exact solution is not possible, and one must resort to various approximations. The starting point for almost all calculations of multielectron atoms is called the **independent-particle approximation**, or **IPA**. This approach is familiar from the classical theory of the solar system, in which one starts by treating the motion of each planet independently, taking account of just the dominant force, the attraction of the sun. Once one has found the orbits of all the planets separately, one can, if desired, improve these approximate orbits by correcting for the small attraction of the planets for one another.

In treating the motion of one planet, it is a very good approximation to ignore the forces of the other planets in comparison with the force of the sun. The corresponding approximation for an atom is much less satisfactory: If, for example, we consider a neutral atom with 20 electrons (calcium), it is true that the single largest force on any one electron is the force of the nucleus, with charge $20e$; but it would be a very poor approximation to consider *only* this force and to ignore completely the repulsion of the 19 other electrons, with a combined charge of $-19e$. What we do instead is to treat the motion of each electron independently, taking account of the force of the nucleus *plus* the force of the *average static distribution of the $Z - 1$ other electrons*. We will refer to the potential energy $U(\mathbf{r})$ of each electron treated in this way as the **IPA potential energy**. The problem is now reduced to finding what the IPA potential energy $U(\mathbf{r})$ should be and then, given $U(\mathbf{r})$, to solving the Schrödinger equation to find the possible wave functions and energies of each electron in the multielectron atom.

To implement this approach quantitatively requires a whole series of successive approximations. One must first make some reasonable guess for the electron wave functions; from these wave functions, one can calculate the charge distribution in the atom and hence the IPA potential-energy function $U(\mathbf{r})$ of each electron. Using this potential-energy function, one can solve the Schrödinger equation for each electron and obtain an improved set of wave functions. Using these improved wave functions, an improved IPA potential-energy function $U(\mathbf{r})$ can be calculated; and so on. This iterative procedure is called the **Hartree-Fock method** and, with the aid of a large computer, can yield quite accurate atomic wave functions and energy levels.

Fortunately, we do not need to go into any details of the Hartree-Fock procedure here. Using simple known properties of the atomic charge distribution, we can get an excellent qualitative picture of the IPA potential-energy function $U(\mathbf{r})$. Using this knowledge of $U(\mathbf{r})$, we can get a good — sometimes even quantitative — understanding of the electron wave functions and hence of atomic structure.

The essential feature of the independent-particle approximation is that each electron can be considered to move independently in the average field of the $Z - 1$ other electrons plus the nucleus. In most atoms it is a good approximation to assume further that the distribution of the $Z - 1$ other electrons is spherically symmetric around the nucleus. (With this additional assumption, the IPA is often called the **central-field approximation**.) We will therefore assume that the charge distribution “seen” by any one electron is spherically symmetric, which means that the IPA potential energy $U(\mathbf{r})$ is independent of θ and ϕ , and can be written as $U(r)$. This greatly simplifies our discussion, for

Section 10.2 • The Independent-Particle Approximation **309**

as we saw in Chapter 8, when $U(r)$ is spherically symmetric, the angular part of the Schrödinger equation always has the same solutions, characterized by the familiar angular-momentum quantum numbers l and m .

The main features of $U(r)$ are easily understood if we recall two properties of spherical charge distributions, both of which follow from Gauss's law. First, if an electron is *outside* a spherical distribution of total charge Q , the electron experiences exactly the same force as if the entire charge Q were concentrated at $r = 0$.

$$F = k \frac{Qe}{r^2} \quad (10.1)$$

Second, if the electron is *inside* a spherical shell of charge, it experiences no force at all from the shell.

If the electron in which we are interested is close enough to the nucleus, it will be inside all the other electrons and will experience the force of the nuclear charge Ze but no force at all from the other electrons; in other words, it feels the full attractive force of the nucleus.

$$F = \frac{Zke^2}{r^2} \quad [r \text{ inside all other electrons}] \quad (10.2)$$

If we now imagine the electron moving outward from the nucleus, it will steadily move outside more and more of the other electrons; thus the force will still be given by (10.1) but with Q equal to the nuclear charge Ze reduced by the charge of those electrons inside the radius r . Eventually, when the electron is outside all of the other $Z - 1$ electrons, Q is given by Ze minus $(Z - 1)e$; that is, $Q = e$ and

$$F = \frac{ke^2}{r^2} \quad [r \text{ outside all other electrons}] \quad (10.3)$$

Therefore, an atomic electron that is outside all its fellow electrons experiences the same force as the one electron in hydrogen.

The potential energy $U(r)$ of the electron is the integral of the force. It follows from this discussion that when r is outside all the other electrons

$$U(r) = -\frac{ke^2}{r} \quad [r \text{ outside all other electrons}] \quad (10.4)$$

On the other hand, as r approaches zero,

$$U(r) \approx -\frac{Zke^2}{r} \quad [\text{as } r \rightarrow 0] \quad (10.5)$$

since, then, the electron of interest is inside all the others.* Between these two regions, $U(r)$ connects these two functions smoothly, as shown qualitatively in Fig. 10.1.

*It follows from (10.2) that $U(r)$ is $-Zke^2/r$ plus some constant. However, as $r \rightarrow 0$, this constant is negligible compared to $-Zke^2/r$, and we have ignored it in (10.5). The corresponding constant in (10.4) is exactly zero since we are defining U to be zero at $r = \infty$.

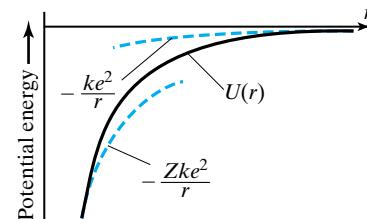
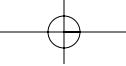


FIGURE 10.1

The IPA potential energy $U(r)$ of an atomic electron in the field of the nucleus plus the average distribution of the other $Z - 1$ electrons. As $r \rightarrow \infty$, U approaches $-ke^2/r$; as $r \rightarrow 0$, U approaches $-Zke^2/r$ as in Eq. (10.5).



310 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

We can express the behavior shown in Fig. 10.1 by writing $U(r)$ as

$$U(r) = -Z_{\text{eff}}(r) \frac{ke^2}{r} \quad (10.6)$$

Here $Z_{\text{eff}}(r)$ gives the **effective charge** “felt” by the electron and depends on r . When r is inside all the other electrons, Z_{eff} approaches the full nuclear charge:

$$Z_{\text{eff}} \approx Z \quad [r \text{ inside all other electrons}] \quad (10.7)$$

As r increases and the nuclear charge is shielded, or *screened*, by more and more of the other electrons, Z_{eff} decreases steadily, until as r moves outside all other electrons, Z_{eff} approaches 1:

$$Z_{\text{eff}} \approx 1 \quad [r \text{ outside all other electrons}]. \quad (10.8)$$

10.3 The IPA Energy Levels

Once the potential energy $U(r)$ is known, our next step is to find the energy levels and wave functions of each electron. The potential energy $U(r)$ is sufficiently like the hydrogen potential energy that we can get a good qualitative understanding of the solutions by analogy with what we already know about hydrogen and hydrogen-like ions.

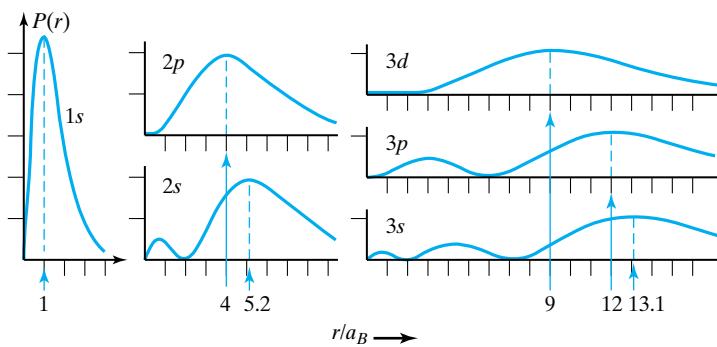
Just as with hydrogen, $U(r)$ depends only on r , and the Schrödinger equation separates. In particular, the two angular equations are exactly the same as for hydrogen. This means that the states of definite energy have angular momentum given by the familiar orbital quantum numbers l and m , and all of the $2l + 1$ different orientations given by $m = l, l - 1, \dots, -l$ have the same energy. Since $U(r)$ does not involve the spin at all, the energy is also the same for both orientations, $m_s = \pm \frac{1}{2}$, of the spin. Thus, each level has a degeneracy of at least $2(2l + 1)$.

As was the case with hydrogen, the solutions of the radial equation are characterized by a principal quantum number n ; and a quantum state is completely specified by the four quantum numbers n, l, m , and m_s . The lowest energy level is $1s$ (that is, $n = 1, l = 0$) and is twofold degenerate because of the two possible orientations of the spin. Just as in hydrogen, the $1s$ wave function is concentrated closer to the nucleus than any other wave functions. This means that in the region where the $1s$ wave function is large, $U(r)$ is close to the hydrogen-like potential energy with $Z_{\text{eff}} \approx Z$. Therefore, the $1s$ wave function approximates that of a hydrogen-like ion with nuclear charge Ze ; the $1s$ energy is close to $-Z^2 E_R$,

$$E_{1s} \approx -Z^2 E_R \quad (10.9)$$

(where E_R denotes the Rydberg energy, 13.6 eV, as usual) and the most probable radius is about a_B/Z (as described in Section 8.10).

Just as with hydrogen, the next energy level has $n = 2$. But here an important difference emerges. In hydrogen the $2s$ and $2p$ states are degenerate, whereas in multielectron atoms the $2s$ states are lower in energy. This difference is easy to understand if we look at the $2s$ and $2p$ radial distributions

**FIGURE 10.2**

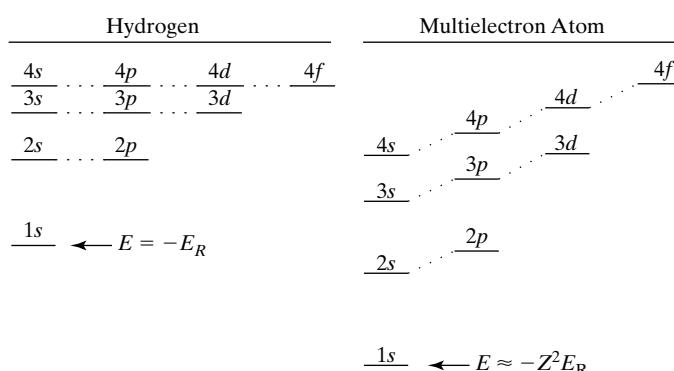
The radial probability distributions for the $n = 1, 2$, and 3 states in hydrogen. The numbers shown are the most probable radii in units of a_B .

shown in Fig. 8.23, which we reproduce here as Fig. 10.2. These two distributions peak at about the same radius, four or five times further out than the $1s$ distribution. This means that the $2s$ and $2p$ wave functions are concentrated in a region where the nuclear charge Ze is screened by any electrons in the $1s$ states, and the $2s$ and $2p$ electrons see an effective charge Z_{eff} which is less than Z . However, the $2s$ distribution (unlike the $2p$) has a secondary maximum much closer in. That is, a small part of the $2s$ distribution penetrates the region where Z_{eff} is close to the full nuclear value $Z_{\text{eff}} \approx Z$. Therefore, on average, a $2s$ electron is more strongly attracted to the nucleus than is a $2p$ electron. This means that the $2s$ electron is more tightly bound and has lower energy.

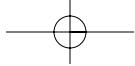
With the $n = 3$ states, there is a similar separation of energies. In Fig. 10.2 it can be seen that both the $3s$ and $3p$ distributions have secondary peaks near $r = 0$, with one of the $3s$ peaks much closer in. Therefore, the $3s$ state penetrates closest to the nucleus and has the lowest energy, the $3p$ is next, and finally the $3d$. This trend is repeated in all higher levels: For each value of n , states with smaller l penetrate closer to the nucleus and are lower in energy. This systematic lowering of the energy for states with lower l is shown schematically in Fig. 10.3.

In many atoms the lowering in energy of the “penetrating orbits” becomes so important that the order of certain levels can be reversed, as compared to hydrogen. This is illustrated in Fig. 10.3, which shows the $4s$ level slightly *below* the $3d$. We will see that such reversals of the order of energy levels have an important effect on the chemical properties of many elements.

For a hydrogen-like ion, we saw that all states with a given n tend to cluster in a spatial shell, with radius roughly equal to the Bohr value $r \approx n^2 a_B/Z$.

**FIGURE 10.3**

Schematic energy-level diagrams for a hydrogen atom and for one of the electrons in a multielectron atom. In hydrogen, all states with the same n are degenerate. In multielectron atoms states with lower l are more tightly bound because they penetrate closer to the nucleus. In many atoms this effect results in the $4s$ level being lower than the $3d$, as shown here.



312 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

This same clustering into spatial shells occurs in multielectron atoms and is, in fact, more pronounced. The $n = 1$ states are closest to the nucleus and feel nearly the full nuclear charge Ze ; therefore, their most probable radius is close to the Bohr value, a_B/Z , for the $n = 1$ state of a hydrogen-like ion with charge Ze . The states with successively higher n are concentrated at progressively larger radii where they feel an effective charge $Z_{\text{eff}}e$ that is progressively *smaller*. The most probable radius for these states is roughly

$$r_{\text{mp}} \approx \frac{n^2 a_B}{Z_{\text{eff}}} \quad (10.10)$$

Since Z_{eff} gets *smaller* as n gets *larger*, the proportionate separation of the spatial shells is even greater in a multielectron atom than it is in hydrogen. Notice that this clustering into spatial shells is according to n , exactly as it is in hydrogen. This contrasts with the energy levels, whose order deviates, as we have seen, from the simple hydrogen ordering.

Occasionally, the IPA energy levels can be calculated easily and with surprising accuracy, as the following example illustrates:

Example 10.1

Wolfgang Pauli
(1900–1958, Austrian)



At the age of 21, Pauli published a review of relativity that is still regarded as a masterpiece. He made many fundamental contributions to quantum physics, including the exclusion principle (1925) for which he won the 1945 Nobel Prize in physics, the neutrino hypothesis (Chapter 17), and work in relativistic quantum field theory. His powerful personality was legendary. He generously credited others with ideas that he originated and detailed in letters to colleagues, but did not publish. Yet he often displayed a biting wit and mercilessly denounced any evidence of sloppy thinking among his peers. Of a paper submitted by a colleague, he said: "This isn't right. This isn't even wrong."

As we will see in Section 10.6, the ground state of lithium ($Z = 3$) has two electrons in the $1s$ level and one in the $2s$. Estimate the energy of the third electron when it is raised to the $3d$ level.

When the outer electron has been raised to the $3d$ level, its wave function is concentrated far outside two inner, $1s$, electrons. Thus, to a very good approximation, we can say that it is moving outside an effective charge of $+e$ (that is, $Z_{\text{eff}} = 1$), and its energy should be just the hydrogenic energy $E_3 = -E_R/Z^2 = -1.512$ eV. This means that the energy to remove the outer electron (in this level) should be 1.512 eV, a prediction that agrees outstandingly well with the observed value of 1.513 eV.

10.4 The Pauli Exclusion Principle

Knowing the possible states of each electron in a multielectron atom, we are now ready to discuss the possible states of the whole atom. Let us consider first the atomic ground states. For these, our problem is to decide how the electrons are to be distributed among their possible states so that the atom as a whole has the minimum possible energy.

One might expect that the ground state of any atom would be found by placing all of its Z electrons into the lowest, $1s$, state; but this is not what is observed to happen. The explanation for what does happen was discovered by the Austrian physicist Wolfgang Pauli, who proposed a new law, now called the **Pauli exclusion principle**. This principle states that*:

*The Pauli principle applies to many other particles besides the electron. For example, it applies to protons and to neutrons and has important consequences for the energy levels of nuclei, as we will see in Chapter 16. In this chapter, however, we are concerned only with electrons.

PAULI EXCLUSION PRINCIPLE

No two electrons in a quantum system
can occupy the same quantum state. **(10.11)**

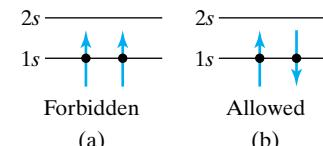
Pauli was led to this law by a study of the states of many atoms, and to this day, the best evidence for the Pauli principle is its success in explaining the diverse properties of all the atoms. There is, however, evidence from many other fields as well. For example, the electrons in a conductor are found to obey the exclusion principle, and many of the observed properties of conductors (conductivity, specific heat, magnetic susceptibility, etc.) depend in a crucial way on the validity of the principle.

To illustrate the exclusion principle and some of the evidence for it, let us consider two simple atoms, helium and lithium. First, let us imagine putting together a helium atom ($Z = 2$) from a helium nucleus and two electrons. If we add one electron to the nucleus, its lowest possible state is the $1s$ state ($n = 1, l = m = 0$), with its spin either up or down ($m_s = \pm \frac{1}{2}$). If we next add the second electron, it too can go into the $1s$ state. But according to the exclusion principle, the two electrons cannot occupy exactly the same quantum state. Since they have the same values of n, l , and m , they must have different values of m_s ; that is, if both electrons are in the $1s$ state, their spins must be antiparallel. This situation is sketched in Fig. 10.4, where part (a) shows two electrons in the $1s$ state with spins parallel, a situation that is never observed; on the other hand, Fig. 10.4(b) shows two electrons in the $1s$ state with spins antiparallel, the situation that *is* observed. The two possibilities shown in Fig. 10.4 would be easily distinguishable, since the first would have a nonzero magnetic moment, while the second has $\mu = 0$. That the helium ground state is always found to have $\mu = 0$ is clear evidence for the exclusion principle.

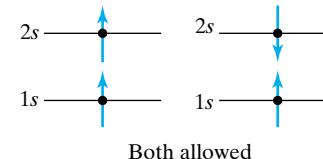
The situation with the excited states of helium is different. For example, there is an excited state with one electron in the $1s$ level and the other in the $2s$ level. In this case the two electrons are certainly in different quantum states, whatever their spin orientations (parallel or antiparallel), as shown in Fig. 10.5. Thus, the Pauli principle does not forbid either of these arrangements, and both are observed, the first with $\mu \neq 0$ and the second with $\mu = 0$.

As a second example, let us imagine putting together a lithium atom ($Z = 3$) from a lithium nucleus and three electrons. When we add the first two electrons, they can both go into the $1s$ level, provided that their spins are antiparallel. But since there are only two possible orientations of the spin, there is now no way in which the third electron can go into the $1s$ level [Fig. 10.6(a)]. The Pauli principle requires that the third electron go into some higher level, the lowest of which is the $2s$. Therefore, the ground state of lithium has to be as shown in Fig. 10.6(b), with the third electron in the $2s$ level and its spin either up or down.

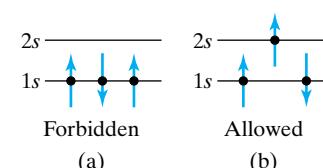
In general, the Pauli principle implies that any s level ($1s, 2s$, etc.) can accommodate two electrons but no more. Levels with higher angular momentum can accommodate more electrons because their degeneracy is larger. For example, any level with $l = 1$ contains six distinct quantum states ($6 = 2 \times 3$ since there are two orientations of S and three orientations of L); therefore, any p level ($l = 1$) can accommodate six electrons, but no more. Similarly, any d level, with $l = 2$, has ten distinct states (2×5) and can accommodate ten electrons, but no more.

**FIGURE 10.4**

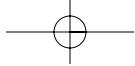
The ground state of helium has both electrons in the $1s$ level. As required by the Pauli principle, their spins have to be in opposite directions.

**FIGURE 10.5**

The lowest excited state of helium has one electron in the $1s$ and one in the $2s$ level. The Pauli principle places no restrictions on the spin orientations in this case.

**FIGURE 10.6**

- (a) It is impossible to put three electrons in the $1s$ level without violating the Pauli principle.
- (b) Therefore, the ground state of lithium has two electrons in the $1s$ level and one in the $2s$ level. The third electron's spin can point either way.



314 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

10.5 Fermions and Bosons; the Origin of the Pauli Principle*

* In this section we describe how the Pauli principle follows from certain symmetry properties of the multiparticle wave function. The ideas described here are of great theoretical importance but will not be used again until Chapter 13, and this section could be omitted on a first reading.

Before exploring further the consequences of the Pauli principle, we take a moment to describe in this section where the principle comes from. One can take the view that the Pauli principle is an observed property of electrons (and some other particles, including protons and neutrons) — a property for which there is overwhelming experimental evidence. If you would like to take this view, then you can safely skip this section for now. Nevertheless, the Pauli principle actually follows from a more fundamental idea — the complete indistinguishability of identical particles in quantum mechanics. This result is interesting in its own right and has a remarkable consequence. All the particles of nature fall into just two categories: First there are the so-called *fermions*, which *do* obey the Pauli principle, and second there are the *bosons*, which do not. We will describe some of the striking differences between these two kinds of particles in Chapter 13.

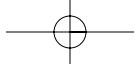
We say that two particles are **identical** if they have all the same intrinsic properties — same mass, same charge, and same spin. Thus any two electrons are identical, but an electron and a proton are certainly not. In classical mechanics, even identical particles are **distinguishable** in the sense that we could, in principle at least, keep track of which is which: Consider, for example, the two electrons in a helium atom. In the classical view each electron follows a definite orbit around the nucleus. Armed with a powerful enough microscope, we could label as number 1 the electron that is on the right at noon today; and then by following their orbits carefully, we could still say tomorrow which is electron 1 and which electron 2. In quantum mechanics this experiment is doomed to failure. We might imagine, in principle at least, measuring the two electrons' positions at noon today and labeling as number 1 the one on the right. But within some fraction of a second their two wave functions will overlap, and when we measure their positions again there is absolutely no way of saying which electron is which. We say that in quantum mechanics two identical particles are **indistinguishable** — we simply cannot say which is which.

Perhaps surprisingly, the indistinguishability of identical particles in quantum mechanics has profound consequences. To describe these, we have to discuss the wave functions for two or more particles. For two spinless particles in one dimension this would have the form $\psi(x_1, x_2)$ where

$$|\psi(x_1, x_2)|^2 dx_1 dx_2 = \begin{cases} \text{probability of finding one particle between } x_1 \\ \text{and } x_1 + dx_1, \text{ and the other between } x_2 \\ \text{and } x_2 + dx_2 \end{cases} \quad (10.12)$$

If the same particles have spin, the wave function would have to have another variable to identify the spin states of the particles. For instance, we might write $\psi = \psi(x_1, m_1, x_2, m_2)$, where

$$|\psi(x_1, m_1, x_2, m_2)|^2 dx_1 dx_2 = \begin{cases} \text{probability of finding one particle} \\ \text{between } x_1 \text{ and } x_1 + dx_1 \text{ with} \\ S_z = m_1\hbar \text{ and the other between } x_2 \\ \text{and } x_2 + dx_2 \text{ with } S_z = m_2\hbar \end{cases} \quad (10.13)$$

Section 10.5 • Fermions and Bosons; the Origin of the Pauli Principle **315**

In three dimensions the variables x_1 and x_2 would be replaced by position vectors \mathbf{r}_1 and \mathbf{r}_2 . To cover all these cases, we will simply write the two-particle wave function as $\psi = \psi(1, 2)$, where the “1” and “2” are shorthand for whatever variables are needed to identify each particle. For instance, in the case of Eq. (10.13), “1” stands for (x_1, m_1) . In most of the discussion that follows, you might want to focus on the simplest case of Eq. (10.12), for which “1” is just short for x_1 and “2” for x_2 .

Suppose first that our two particles are not identical; for example, $\psi(1, 2)$ could be a state of a high-energy electron and a low-energy proton. In this case the wave function $\psi(2, 1)$, in which the roles of the two particles are reversed, would represent a high-energy proton and a low-energy electron — an entirely distinct situation. But suppose instead that $\psi(1, 2)$ were the wave function for two *identical* particles, two electrons for instance. The indistinguishability of identical particles requires that the states represented by $\psi(1, 2)$ and $\psi(2, 1)$ (with the roles of 1 and 2 reversed) must be physically indistinguishable. It makes no difference which is particle 1 and which is particle 2. In particular, the probability densities associated with $\psi(1, 2)$ and $\psi(2, 1)$ must be the same:

$$|\psi(2, 1)|^2 = |\psi(1, 2)|^2 \quad (10.14)$$

It turns out that there are only two ways in which this indistinguishability requirement can be met: For a given kind of particle (electron or photon, for instance), *either*

$$\psi(2, 1) = +\psi(1, 2) \quad (10.15)$$

for all two-particle states, or

$$\psi(2, 1) = -\psi(1, 2) \quad (10.16)$$

for all two-particle states.* Wave functions that satisfy (10.15) are said to be **symmetric** under particle exchange; particles whose multiparticle wave functions are symmetric in this way include photons and pions and are called **bosons**, after the Indian physicist Satyendranath Bose. Wave functions that satisfy (10.16) are said to be **antisymmetric** under particle exchange; particles whose multiparticle wave functions are antisymmetric include electrons, protons, and neutrons and are called **fermions**, after the Italian-American physicist Enrico Fermi.

It is found experimentally that all bosons have integer spin, $s = 0, 1, 2, \dots$, whereas all fermions have half-odd-integer spin,† $s = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$. This connection often lets one decide quickly whether a given particle is a fermion or boson. For example, electrons, protons, and neutrons all have spin half, so have to be fermions; photons have spin 1, so have to be bosons. We will describe several of the dramatic differences between bosons and fermions in Chapter 13. Here we will just describe how the requirement (10.16) implies that fermions obey the Pauli exclusion principle.

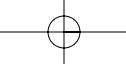
*The proof that (10.15) and (10.16) are the only two possibilities is actually quite simple, though it requires a little more knowledge of quantum mechanics than we have yet developed. It is one of the fundamental postulates of quantum mechanics that if two wave functions represent the same physical state, one must be a constant multiple of the other. Therefore, in our case $\psi(2, 1) = k\psi(1, 2)$. Interchanging the particles again, we find that $\psi(1, 2) = k\psi(2, 1) = k^2\psi(1, 2)$, which implies that $k^2 = 1$ or $k = \pm 1$.

† In relativistic quantum field theory, one can in fact *prove* this connection.

**Satyendranath
Bose**
(1894–1974, Indian)



Bose was born and educated in Calcutta, India. In a paper written in 1924 he derived the Planck formula for blackbody radiation by treating the photons as what we would now call bosons. This paper drew the attention of Einstein and secured an invitation for Bose to visit Europe, where he met Einstein, de Broglie, Born, and others. Einstein extended Bose's ideas, and the rules that govern bosons are now called Bose-Einstein statistics. We will see some of the dramatic consequences of these ideas in Chapter 13.



316 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

Consider a two-particle wave function that happens to be a simple product

$$\psi(1, 2) = \phi(1)\chi(2) \quad (10.17)$$

Suppose first the two particles concerned are nonidentical, an electron and a proton, for instance. Then the wave function (10.17) represents a state in which the electron occupies the state ϕ and the proton occupies the state χ . The function $\psi(2, 1) = \phi(2)\chi(1)$ represents the completely distinct state with the proton in state ϕ and the electron in state χ . Suppose, however, the two particles are identical fermions (two electrons, for example). In this case, their wave function must satisfy the antisymmetry requirement (10.16), which (10.17) does not (unless one of the functions ϕ or χ is identically zero). The only way to reconcile (10.17) with (10.16) is to use the antisymmetric combination

$$\psi(1, 2) = \phi(1)\chi(2) - \chi(1)\phi(2) \quad (10.18)$$

[Notice that this automatically satisfies $\psi(2, 1) = -\psi(1, 2)$. Note also that as it stands, this wave function is generally not normalized; this requirement is easily taken care of, but need not concern us here.] A wave function of the form (10.18) represents two identical fermions, one of which is in state ϕ and the one in state χ (though we can't say which is which). Notice that this actually makes good sense: We can view (10.18) as an equal mixture of two states, one with particle 1 in state ϕ and particle 2 in state χ and the other with the two particles reversed. Given the indistinguishability of the two particles, this is a very natural compromise.

The “antisymmetrized” wave function (10.18) is the only way to construct a state of two identical fermions with one in the state ϕ and one in the state χ . But if $\phi = \chi$, then (10.18) is identically zero. Therefore, there is no way to construct a state in which two identical fermions occupy the same one-particle state, and this is the Pauli principle.

If our two identical particles were bosons, then the same argument shows that the wave function with one particle in the state ϕ and one in the state χ must be the “symmetrized” combination

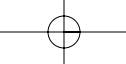
$$\psi(1, 2) = \phi(1)\chi(2) + \chi(1)\phi(2) \quad (10.19)$$

which gives no trouble if $\phi = \chi$. Thus, two identical bosons *can* occupy the same one-particle state, and bosons do not have to satisfy the Pauli principle. In fact, as we will see in Chapter 13, while two identical fermions cannot occupy the same state, there is a sense in which identical bosons actually *prefer* to occupy the same state.

10.6 Ground States of the First Few Elements

Let us return to the consequences of the Pauli principle for multielectron atoms. To determine the ground state of an atom, we have only to assign its Z electrons to the lowest individual energy levels consistent with the Pauli principle (that no two electrons occupy the same quantum state). In this section we use this procedure to find the ground states of the lightest few atoms, starting with hydrogen ($Z = 1$) and going as far as sodium ($Z = 11$).

The ground state of hydrogen has its one electron in the $1s$ level, with its spin pointing either way. The energy is $E = -E_R = -13.6 \text{ eV}$, which means

Section 10.6 • Ground States of the First Few Elements **317**

that the energy needed to remove the electron — the **ionization energy** — is 13.6 eV.

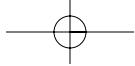
Moving on to helium, we already saw that the ground state has both electrons in the 1s level, with their spins antiparallel. Because of the greater nuclear charge ($Z = 2$), the 1s level of helium is much lower in energy than that of hydrogen. If we write the energy of either electron as $-Z_{\text{eff}}^2 E_R$, then Z_{eff} will not be equal to the full nuclear charge, 2, since each electron is screened by the other. Nevertheless, Z_{eff} should be appreciably more than 1 — perhaps around 1.5. Thus, we would expect that helium should be significantly more tightly bound than hydrogen. This prediction is well borne out by experiment: The first ionization energy of helium (the energy to remove one electron) is found to be 24.6 eV, nearly twice that of hydrogen; since the ionization energy is proportional to Z_{eff}^2 , this implies that Z_{eff} is close to $\sqrt{2} \approx 1.4$.

Another measure of an atom's stability is its first excitation energy, the energy to lift it to its first excited state. In both H and He this involves lifting one electron from the 1s to the 2s level. In helium this should require a larger energy by a factor of roughly Z_{eff}^2 , the same ratio as for the ionization energies. This, too, is confirmed by experiment: The first excitation energy of He is 19.8 eV, compared to 10.2 eV in H.

The ionization and excitation energies of an atom are important indicators of the atom's stability. On both of these counts, helium is about twice as stable as hydrogen. In fact, helium has the largest ionization and excitation energies of *any* atom. Since high stability tends to imply low chemical activity, we might guess that helium should be chemically inactive; and this proves to be the case. Helium is one of the six **noble, or inert, gases**, which show almost no chemical activity at all, form no really stable compounds, and can bind together into liquid or solid form only at relatively low temperatures.

Another important difference between the hydrogen and helium atoms concerns their sizes. We have seen that the wave functions of a hydrogen-like ion are scaled inward by a factor of $1/Z$, compared to the corresponding wave functions of hydrogen. Therefore, the radius of the 1s wave function of helium should be about $1/Z_{\text{eff}}$ times that of hydrogen, and the He atom should therefore be roughly two-thirds the size of the H atom. This prediction also is correct. The precise value of the atomic radius depends on how one chooses to define it, but representative values are 0.08 nm for H and 0.05 nm for He.

The differences in energy and radius between hydrogen and helium reflect the larger nuclear charge ($Z = 2$) of helium. When we consider the lithium atom ($Z = 3$), we encounter a new kind of difference, due to the Pauli exclusion principle. Let us consider first an electron in the 1s level of Li. Because of the greater nuclear charge, this 1s electron is more tightly bound and concentrated at a smaller radius than a corresponding electron in either He or H. However, the Pauli principle allows only two of lithium's three electrons to occupy the 1s level; the third electron must occupy the 2s level and is much less tightly bound. In fact, we can easily estimate the binding energy of this outermost electron: Because it is outside the two other electrons, it sees an effective charge of order $Z_{\text{eff}} = 1$, about the same as for the one electron in hydrogen. Therefore, since it is in the $n = 2$ level, it should have about the same ionization energy as an $n = 2$ electron in hydrogen, 3.4 eV. This estimate agrees reasonably with lithium's observed ionization energy of 5.4 eV. (That the actual value, 5.4 eV, is a bit larger than our estimate of 3.4 eV shows that the outer electron is not perfectly shielded by the inner two and sees an effective charge somewhat greater than $Z_{\text{eff}} = 1$.) This ionization energy is the fifth smallest of any stable atom and means that the lithium atom can easily lose its



318 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

outermost electron. This is the main reason why lithium is so chemically active, as we describe in Section 12.2.

Because the outer electron of lithium is in the $n = 2$ level, the Li atom should have a much larger radius than either He or H. This prediction is confirmed by the data in Table 10.1, which shows the ionization energies and radii of the first four atoms, ${}_1\text{H}$, ${}_2\text{He}$, ${}_3\text{Li}$, and ${}_4\text{Be}$. (When convenient, we indicate the atomic number, Z , by a *subscript* on the left of the atomic symbol — not to be confused with the mass number, A , which is sometimes shown as a *superscript* on the left.)

TABLE 10.1

Ionization energies and radii of the first four atoms. Atomic numbers Z are shown as subscripts on the left of chemical symbols. The energy levels are not to scale, since corresponding levels get deeper as Z increases.

	${}_1\text{H}$	${}_2\text{He}$	${}_3\text{Li}$	${}_4\text{Be}$
Ionization energy (eV):	13.6	24.6	5.4	9.3
Radius (nm):	0.08	0.05	0.20	0.14
Occupancy of energy levels:	$2s$ ———	———	\uparrow \downarrow	\uparrow \downarrow \downarrow
	$1s$ \uparrow \downarrow	\uparrow \downarrow	\uparrow \downarrow	\uparrow \downarrow

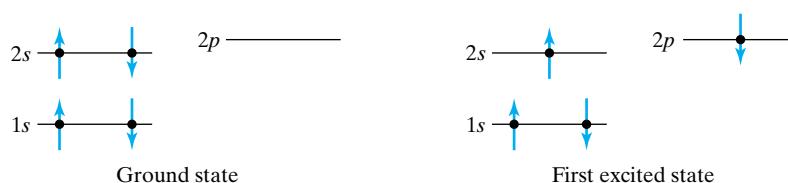
In beryllium ($Z = 4$) the fourth electron can join the third electron in the $2s$ level, provided that their spins are antiparallel. Because of the larger nuclear charge, this level is more tightly bound and its radius smaller than in ${}_3\text{Li}$. Therefore, the ${}_4\text{Be}$ atom should have a larger ionization energy and a smaller radius, as the data in Table 10.1 confirm.

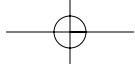
To some small extent, the ${}_4\text{Be}$ atom with its filled $2s$ level is similar to the ${}_2\text{He}$ atom with its filled $1s$ level. But the differences are more important than the similarities. In particular, the ${}_2\text{He}$ atom is not only hard to ionize, it is also hard to excite, 19.8 eV being needed to lift one of the electrons from the $1s$ to the $2s$ level. Excitation of the ${}_4\text{Be}$ atom requires only that one of the $2s$ electrons be lifted to the nearby $2p$ level, just 2.7 eV higher (see Fig. 10.7). This means that one of the electrons in Be can easily move to the higher level. As we will see in Chapter 12 (see especially Problem 12.30 XXX), this allows Be to bond to other atoms. For this reason, Be, unlike He, is chemically active and forms a number of compounds.

After beryllium (${}_4\text{Be}$) comes boron (${}_5\text{B}$). The first four electrons of ${}_5\text{B}$ go into the $1s$ and $2s$ levels, just as with ${}_4\text{Be}$. But as required by the Pauli principle, the last electron of ${}_5\text{B}$ must occupy the $2p$ level. Therefore, in moving from ${}_4\text{Be}$ to ${}_5\text{B}$, we see two competing effects: The increase in Z causes any given

FIGURE 10.7

Excitation of beryllium ($Z = 4$) requires only 2.7 eV to lift one of the $2s$ electrons to the nearby $2p$ level. In the excited state the spins of the $2s$ and $2p$ electrons can point either way.



Section 10.6 • Ground States of the First Few Elements **319**

level to be somewhat more tightly bound, but the final electron has to occupy a level that is slightly higher and hence somewhat less tightly bound. As far as ionization energy is concerned, the second effect wins: The ionization energy of ${}_5\text{B}$ is 8.3 eV, just a little less than the 9.3 eV of ${}_4\text{Be}$. On the other hand, the radius of ${}_5\text{B}$ is less than that of ${}_4\text{Be}$ and continues the trend of shrinking radii with increasing Z . In neither case is the difference large.

The six elements after ${}_4\text{Be}$ are

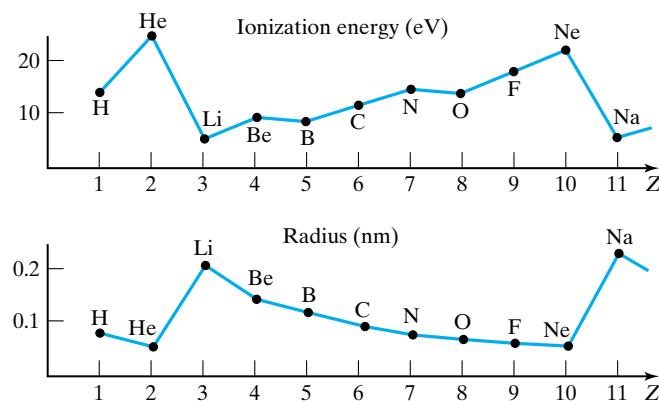
Element:	boron	carbon	nitrogen	oxygen	fluorine	neon
Symbol:	${}_{ 5}\text{B}$	${}_{ 6}\text{C}$	${}_{ 7}\text{N}$	${}_{ 8}\text{O}$	${}_{ 9}\text{F}$	${}_{ 10}\text{Ne}$

In all of these atoms the first four electrons fill the $1s$ and $2s$ levels. Since the $2p$ level can hold six electrons, the remaining electrons all go into the $2p$ level. Thus, as we move from ${}_{ 5}\text{B}$ to ${}_{ 10}\text{Ne}$, the additional electron of each succeeding atom goes into the same level, and the main differences should be due to the increasing nuclear charge Z ; in particular, the ionization energy should increase and the radius should decrease. These trends show up clearly in Fig. 10.8, in which the ionization energies and atomic radii are plotted as functions of Z for all the first 11 atoms. With one small exception, the two graphs change steadily in the expected directions (ionization energy increasing, radius decreasing) as Z increases from 5 to 10. The one exception is the small drop in ionization energy as one moves from ${}_{ 7}\text{N}$ to ${}_{ 8}\text{O}$; we will return to this anomaly later.

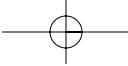
When we reach ${}_{ 10}\text{Ne}$, the $2p$ level has its full allotment of six electrons. Therefore, when we move on to sodium, ${}_{ 11}\text{Na}$, the last electron must go into the next, and much higher, level — the $3s$ level. This reverses all of the trends set by the last eight atoms: The ionization energy drops abruptly, and the radius increases abruptly, as is shown clearly in Fig. 10.8.

Both of the graphs in Fig. 10.8 suggest a parallel between ${}_{ 3}\text{Li}$ and ${}_{ 11}\text{Na}$. Both atoms have unusually low ionization energies and unusually large radii. The low ionization energies mean that both atoms can easily lose one electron. This allows lithium and sodium to combine with other atoms to form many different chemical compounds and is the reason why both atoms are chemically so active, as we describe in Chapter 12.

The similarity between ${}_{ 3}\text{Li}$ and ${}_{ 11}\text{Na}$ is an example of the *periodic* behavior of the elements: As we consider elements with successively higher Z , chemical similarities recur at certain regular, or periodic, intervals. Another example of this periodicity is the pair of elements ${}_{ 2}\text{He}$ and ${}_{ 10}\text{Ne}$. Both are very stable (large ionization and excitation energies) and very small in size. In Chapter 8 we mentioned that the word “shell” is often used for a group of

**FIGURE 10.8**

The ionization energies and atomic radii of the first 11 elements.



320 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

energy levels that are close to one another and well separated from any others. From the first graph in Fig. 10.8, it is clear that the $1s$ level should be considered as one shell by itself, and $2s$ and $2p$ together as another. For this reason, helium and neon are called closed-shell atoms. We will see that there are six closed-shell atoms in all and that they are the six noble gases. In the same way, lithium and sodium can be described as being **closed-shell-plus-one** and are the first of six such elements, called the **alkali metals**.

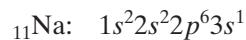
Just before the stable $_{10}\text{Ne}$ is fluorine, $_{9}\text{F}$. The ionization energy of fluorine is 17.4 eV, which is the third largest ionization energy of any atom. One might, therefore, imagine that fluorine would be chemically inactive, but such is definitely not the case. Fluorine is one of the most active of all the elements. The reason for this activity is that the fluorine atom is **closed-shell-minus-one** since its $2p$ level is one short of the full complement of six electrons. Because of the large nuclear charge, the $2p$ level is very well bound (as the large ionization energy testifies). In fact, the $2p$ level of F is so well bound that it can bind an extra, sixth electron. That is, the negative ion, F^- , is stable, with the extra electron just filling the $2p$ level. The tendency of an atom to bind an extra electron is measured by its **electron affinity**. This is defined as the energy released when the atom captures an extra electron and forms a negative ion (or, equivalently, the energy needed to remove one electron from the negative ion). The electron affinity of fluorine is 3.4 eV, the third largest for any element. As we discuss in Chapter 12, it is because of its ability to bind an extra electron that fluorine is so active.

10.7 The Remaining Elements

In Section 10.6 we examined the ground states of the first 11 elements. In this section we sketch a similar analysis of some of the remaining 90 or so elements. This will emphasize what was already becoming apparent. Because of the Pauli principle, the properties of atoms do not vary smoothly and uniformly as functions of Z . Rather, as we examine atoms with successively more electrons, their properties vary more or less smoothly as long as each extra electron can be accommodated in the same shell; but each time a shell is filled and a new shell comes into play, there is an abrupt change in the properties, reversing the previous smooth trends. As we saw in Section 10.6, this leads to the periodic occurrence of atoms with similar physical and chemical properties.

To find the ground state of an atom, we must assign the Z electrons to the lowest levels consistent with the Pauli principle. For $_{1}\text{H}$ and $_{2}\text{He}$ the electrons go into the $1s$ level. For $_{3}\text{Li}$ through $_{10}\text{Ne}$, the $1s$ level is full, and the outer, or **valence**, electrons go into the $2s$ and then $2p$ levels. Similarly, when we move on to the elements 11 through 18 (sodium through argon), the $1s$, $2s$, and $2p$ levels are all full, and the valence electrons go into the $3s$ and then $3p$ levels.

Perhaps the most descriptive way to show the assignment of electrons to energy levels is with energy-level diagrams like those in Fig. 10.7. Unfortunately, these diagrams become increasingly cumbersome as we discuss atoms with more electrons. A more compact way to show the same information is to give the **electron configuration**, which is just a list of the occupied levels, each with a superscript to indicate the number of electrons in it. For example, the electron configuration of the sodium ground state is



With this notation, the ground states of the first 18 elements are as shown in Table 10.2.

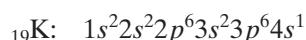
TABLE 10.2

Electron configurations of the ground states of the first 18 elements.

First Shell	Second Shell	Third Shell
${}_1\text{H} : 1s^1$	${}_3\text{Li} : 1s^2 2s^1$	${}_{11}\text{Na} : 1s^2 2s^2 2p^6 3s^1$
${}_2\text{He} : 1s^2$	${}_4\text{Be} : 1s^2 2s^2$	${}_{12}\text{Mg} : 1s^2 2s^2 2p^6 3s^2$
	${}_5\text{B} : 1s^2 2s^2 2p^1$	${}_{13}\text{Al} : 1s^2 2s^2 2p^6 3s^2 3p^1$
	${}_6\text{C} : 1s^2 2s^2 2p^2$	${}_{14}\text{S} : 1s^2 2s^2 2p^6 3s^2 3p^2$
	${}_7\text{N} : 1s^2 2s^2 2p^3$	${}_{15}\text{P} : 1s^2 2s^2 2p^6 3s^2 3p^3$
	${}_8\text{O} : 1s^2 2s^2 2p^4$	${}_{16}\text{S} : 1s^2 2s^2 2p^6 3s^2 3p^4$
	${}_9\text{F} : 1s^2 2s^2 2p^5$	${}_{17}\text{Cl} : 1s^2 2s^2 2p^6 3s^2 3p^5$
	${}_{10}\text{Ne} : 1s^2 2s^2 2p^6$	${}_{18}\text{Ar} : 1s^2 2s^2 2p^6 3s^2 3p^6$

The properties of elements 11 to 18 closely parallel those of elements 3 to 10. As we have already noted, ${}_{11}\text{Na}$ and ${}_3\text{Li}$ are both easily ionized and are chemically very active. As one moves from $Z = 11$ to $Z = 18$ and the $3s$ and $3p$ levels fill, the ionization energies increase (with two small exceptions) and the atomic radii decrease, just as occurred between $Z = 3$ and $Z = 10$. At $Z = 18$ the $3p$ level is completely full and ${}_{18}\text{Ar}$, like ${}_{10}\text{Ne}$, is very stable and chemically inert. Just before ${}_{18}\text{Ar}$ is chlorine (${}_{17}\text{Cl}$), which, like ${}_9\text{F}$, is able to accept an extra electron into the one vacancy in its outer p level and is therefore chemically active.

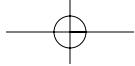
When we move beyond ${}_{18}\text{Ar}$ to ${}_{19}\text{K}$ (potassium), the story becomes more complicated. One might expect that the next level occupied would be the $3d$ level. In fact, however, the tendency for levels with low angular momentum (the “penetrating orbits”) to have lower energy causes the $4s$ level to be slightly lower than the $3d$, as discussed in Section 10.3. Therefore, the configuration of the ground state of ${}_{19}\text{K}$ is



The order in which the energy levels fill is shown in Fig. 10.9. In this picture we have also shown the grouping of the energy levels into **shells** containing levels that are close to one another but well separated from other levels. (Note that in this context an individual level within a shell is sometimes called a **subshell**.) We see, for example, that the $3s$ and $3p$ levels are close together, but that the gap from $3p$ to $4s$ is large. Thus $3s$ and $3p$ form a shell by themselves, just like $2s$ and $2p$. Therefore, ${}_{18}\text{Ar}$ is a closed-shell atom, like ${}_{10}\text{Ne}$ and ${}_{2}\text{He}$, and is the third of the noble gases. The next element, potassium (${}_{19}\text{K}$), is a *closed-shell-plus-one* atom, with low ionization energy, and is one of the alkali metals, similar to its two predecessors, ${}_{11}\text{Na}$ and ${}_3\text{Li}$.

It can be seen in Fig. 10.9 that, as it happens, the lowest level in each shell is always an s level and the highest is a p level (except in the first shell, which has only the $1s$ level). Thus all the closed-shell-plus-one atoms have a single s electron outside a closed-shell core; all closed-shell-minus-one atoms (often called **halogens**) have one vacancy, or *hole*, in the p level of an otherwise filled shell.

AU:
delete?



322 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

FIGURE 10.9

Schematic diagram showing the order in which levels are occupied as one considers atoms with successively higher Z . This is not the energy-level diagram for any one atom; it just gives the order in which levels are occupied as Z increases. The shaded rectangles indicate the groupings of nearby levels into energy shells. The figure to the right of each level or shell gives the number of electrons that can be accommodated; the figures on the far right are the atomic numbers Z of the closed-shell atoms. Note that some levels are too close to be ordered unambiguously; for example, $5d$ is partially occupied before $4f$, but $4f$ is completely filled before $5d$. The exact sequence of occupancy can be found from the electron configurations shown inside the back cover.

7th shell	$\underline{7s} \quad 2$	$\underline{\overline{7p}} \quad 6$	$\underline{\overline{6d}} \quad 10$	$\underline{\overline{5f}} \quad 14$	32
6th	$\underline{6s} \quad 2$	$\underline{\overline{6p}} \quad 6$	$\underline{\overline{5d}} \quad 10$	$\underline{\overline{4f}} \quad 14$	$Z = 86$ 32
5th	$\underline{5s} \quad 2$	$\underline{\overline{5p}} \quad 6$	$\underline{\overline{4d}} \quad 10$		54 18
4th	$\underline{4s} \quad 2$	$\underline{\overline{4p}} \quad 6$	$\underline{\overline{3d}} \quad 10$		36 18
3rd	$\underline{3s} \quad 2$	$\underline{\overline{3p}} \quad 6$			18
2nd	$\underline{2s} \quad 2$	$\underline{\overline{2p}} \quad 6$			10
1st	$\underline{1s} \quad 2$				2

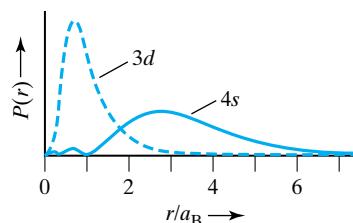


FIGURE 10.10

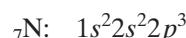
Typical radial probability distributions for a $3d$ and $4s$ electron in the transition elements $_{21}\text{Sc}$ through $_{30}\text{Zn}$. Note how the $4s$ distribution peaks more than four times farther out than the $3d$. For this reason, it is the $4s$ electrons that determine the chemical properties of the transition metals.

After $_{19}\text{K}$ comes calcium ($_{20}\text{Ca}$), whose $4s$ level is full. Then with scandium ($_{21}\text{Sc}$), the $3d$ level begins to fill. Since any d level can hold 10 electrons, the filling of the $3d$ level takes us from $_{21}\text{Sc}$ through $_{30}\text{Zn}$ (zinc). These 10 elements are called **transition metals** and are alike in several ways. This is because the $3d$ wave functions peak at a much smaller radius than the $4s$ functions (Fig. 10.10). Thus, although it is the $3d$ level that is filling in this sequence of elements, it is the $4s$ electrons that are farthest out. Since it is the outer electrons that can interact with other atoms, all of the transition metals from $_{21}\text{Sc}$ to $_{30}\text{Zn}$ have similar chemical properties.

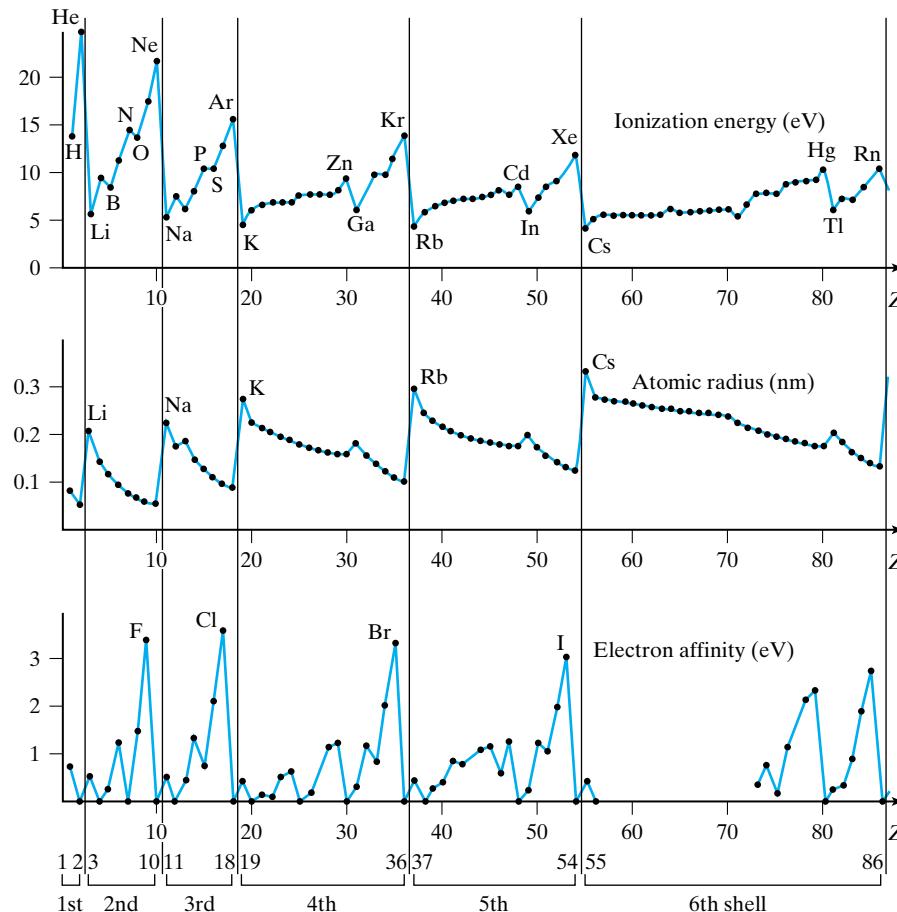
Once the $3d$ level is full at $_{30}\text{Zn}$, the $4p$ level starts to fill, and the sequence of six elements from $_{31}\text{Ga}$ (gallium) to $_{36}\text{Kr}$ (krypton) is analogous to the sequences from $_{13}\text{Al}$ to $_{18}\text{Ar}$ and from $_{5}\text{B}$ to $_{10}\text{Ne}$, in which the $3p$ and $2p$ levels were filling. In particular, $_{36}\text{Kr}$ is a closed-shell atom and is the fourth of the noble gases.

Beyond $_{36}\text{Kr}$, the fifth, sixth, and, finally, seventh shells are occupied, but since there are no really important new features, we need not go into details. The periodic recurrence of similar atomic properties can be clearly seen in plots of a variety of properties as functions of Z . In Fig. 10.11 we show ionization energies, atomic radii, and electron affinities for $Z = 1$ through 86. (The first two graphs are extensions of those shown in Fig. 10.8.) Notice how the ionization energies have pronounced maxima at the closed-shell noble gases and minima at the closed-shell-plus-one alkali metals. The ionization energies have smaller maxima at several other points, and most of these are easily explained. For example, the small peak at $_{4}\text{Be}$ is because the $2s$ level (or subshell) is filled; that at $_{30}\text{Zn}$ is because the $3d$ level is filled.

The small peak in the ionization energy at $_{7}\text{N}$ is a little harder to explain and is due to a subtle but important change when we move from $_{7}\text{N}$ to $_{8}\text{O}$. The configuration of nitrogen is



with three electrons in the $2p$ level. The six distinct quantum states of the $2p$ level arise from the three possible wave functions (the orbitals $2p_x$, $2p_y$, and

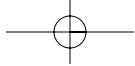
**FIGURE 10.11**

Ionization energy, atomic radius, and electron affinity as functions of atomic number Z . The vertical lines separate complete shells. The electron affinities for elements 57 through 72 have not been measured.

$2p_z$, shown in Fig. 8.21*) each with two possible spin orientations. Therefore, the three outer electrons in nitrogen can occupy three distinct orbitals. This arrangement keeps the electrons well separated, reducing their electrostatic repulsion, and makes the nitrogen atom relatively stable, with a comparatively large ionization energy. In oxygen the fourth $2p$ electron must occupy the same orbital as one of the others (with opposite spin, of course). Because the distributions of these two electrons overlap strongly, their electrostatic repulsion is relatively large, and the ionization energy of oxygen is a little less. It is for this reason that the plot of ionization energy in Fig. 10.11 has the small drop after $_7\text{N}$. For the same reason, there is a small drop in ionization energy after $_{15}\text{P}$ also visible in Fig. 10.11.

The graph of atomic radius against Z shows the expected trends, with R falling to a minimum at each of the noble gases and jumping abruptly to a maximum at each closed-shell-plus-one atom. The downward slope from $_{21}\text{Sc}$ to $_{30}\text{Zn}$ is very gentle because the last electrons of these transition elements go into the $3d$ level, which is concentrated at a smaller radius than the occupied $4s$ level. Thus the atomic radius is determined by the $4s$ electrons and changes little from $Z = 21$ to 30 (from 0.21 to 0.15 nm). The even gentler slope from $Z = 57$ to 71 is a similar effect: With these *inner transition* elements, the level

*One can arrive at the same conclusion using the wave functions with $m = 1, 0, -1$, but the argument is a little more complicated.



324 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

being filled is the $4f$ level, which has *much* smaller radius than the occupied $6s$ level; thus the atomic radius is almost completely determined by the $6s$ radius and changes very little (from 0.27 to 0.22 nm).

The third graph in Fig. 10.11 shows the electron affinity. As expected, this shows sharp maxima at the closed-shell-minus-one atoms (the halogens, $_{9}\text{F}$, $_{17}\text{Cl}$, etc.) and drops abruptly to zero at the closed shells ($_{2}\text{He}$, $_{10}\text{Ne}$, etc.). It is also zero at those atoms with closed subshells, such as $_{4}\text{Be}(1s^22s^2)$ and $_{12}\text{Mg}(1s^22s^22p^63s^2)$. Some atoms with *half-filled* subshells also have zero electron affinity. For example, an eighth electron in $_{7}\text{N}$ would have to occupy the same orbital, $2p_x$, $2p_y$, or $2p_z$, as one of the other $2p$ electrons, and the resulting Coulomb repulsion would raise its energy so high that it would not be bound.

To conclude this section, we mention one more atomic property that can be predicted from our knowledge of the electron states — the angular momentum of the whole atom. For many atoms, calculation of the total angular momentum is difficult and well beyond our present discussion. However, there are several cases that are quite straightforward. This is because the total angular momentum of all the electrons in any filled level is zero. To understand this useful result, we have only to note that in a filled level there are equal numbers of electrons with spin up and spin down; therefore, the total spin $\sum \mathbf{S}$ is certainly zero. Further, for each electron with a given value of L_z , there is another with exactly the opposite value (since all the values $m = l, l-1, \dots, -l$ are occupied); therefore, the total orbital angular momentum $\sum \mathbf{L}$ is also zero. It follows that the total angular momentum of any closed-shell or closed-subshell atom must be zero. For example, the total angular momenta of $_{2}\text{He}$ and $_{10}\text{Ne}$, and of $_{4}\text{Be}$ and $_{30}\text{Zn}$, should be and are zero.

The total angular momenta of the closed-shell-plus-one alkali atoms are also easy to predict. These atoms all contain various filled levels, with zero total angular momenta, plus a single electron in an s level ($l = 0$). Therefore, the total angular momentum of any alkali atom ($_{3}\text{Li}$, $_{11}\text{Na}$, etc.) is equal to the spin angular momentum ($s = \frac{1}{2}$) of its one valence electron.

One can also predict the angular momentum of the closed-shell-minus-one atoms. These all consist of filled levels, with zero total angular momentum, plus a p level containing five of its six possible electrons. We know that the *addition of one more p* electron (with $l = 1$) would produce a filled level, with total $l = 0$. The addition of one vector to another can give zero only if the two vectors have the same magnitude. It follows that the total orbital angular momentum of the original five electrons must also have total $l = 1$. A similar argument shows that the total spin of the five electrons must be $s = \frac{1}{2}$. Therefore, the closed-shell-minus-one atoms must have total $l = 1$ and $s = \frac{1}{2}$, and this is what is observed. The resulting *total* angular momentum depends on the relative orientation of \mathbf{L} and \mathbf{S} ; this, also, can be predicted, but would take us too far afield at present.

10.8 The Periodic Table

The periodic recurrence of atoms with similar properties had been noticed long before the discovery of quantum mechanics and the understanding of atomic structure. In 1869, a German chemist, Lothar Meyer, had plotted several different properties of the elements as functions of their atomic masses. In particular, he had noticed that his plot of atomic volume (which was very

similar to the plot of atomic radius in Fig. 10.11) was divided by five sharp peaks into six distinct sections, or periods, corresponding (as we now know) to the successive filling of the first six shells.

In the same year, 1869, the Russian chemist Dmitri Mendeleev had proposed his **periodic table**, an ingenious array in which the elements are arranged so as to highlight their periodic properties. In most modern periodic tables, the elements are arranged in order of increasing atomic number, placed from left to right in rows, with a new row starting each time a shell is closed. Within each row, the elements are placed so that atoms with similar properties are contained in the same vertical column of the table.

One popular arrangement of the entire table is shown in Fig. 10.12. The horizontal rows, or **periods**, contain different numbers of elements, corresponding to the different capacities of the various shells. The first period contains just two elements, H and He; the next two periods contain eight elements each; and the next two contain 18 each. The sixth period contains 32 elements, from $Z = 55$ to 86, but to save space, the part of this period from $Z = 57$ to 71 is usually detached and arranged in a separate row below. For the same reason, the corresponding elements (from $Z = 89$ to 103) in the seventh period are placed in a second additional row.

The vertical columns, or **groups**, contain elements with corresponding arrangements of their valence electrons. The leftmost group, often called group I, contains the closed-shell-plus-one atoms, each with a single s electron outside a closed core. The next column, group II, has the closed-shell-plus-two atoms, each with two s electrons outside a closed core. On the right side of the table are the six columns, groups III to VIII, whose atoms have a filled or partially

Dmitri Mendeleev
(1834–1907, Russian)

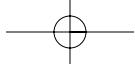


Several chemists had recognized that when arranged by atomic mass, the elements show periodic behavior, but the Russian chemist Mendeleev was unique in his recognition that certain elements were clearly missing from this scheme. When three of the missing elements were discovered with exactly the properties he had predicted, his periodic table gained international acceptance.

Group I		II		Transition elements														VIII																											
Period 1 →		→																→		2 He																									
1 H		4 Be																2 He																											
2 →	3 Li	2 →	4 Be															5 B	6 C	7 N	8 O	9 F	10 Ne																						
3 →	11 Na	3 →	12 Mg															13 Al	14 Si	15 P	16 S	17 Cl	18 Ar																						
4 →	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr																											
5 →	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe																											
6 →	55 Cs	56 Ba	57–71 *	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn																											
7 →	87 Fr	88 Ra	89–103 †	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt																																				
Inner transition elements																																													
*Lanthanides →																																													
†Actinides →																																													
57 La 58 Ce 59 Pr 60 Nd 61 Pm 62 Sm 63 Eu 64 Gd 65 Tb 66 Dy 67 Ho 68 Er 69 Tm 70 Yb 71 Lu																																													
89 Ac 90 Th 91 Pa 92 U 93 Np 94 Pu 95 Am 96 Cm 97 Bk 98 Cf 99 Es 100 Fm 101 Md 102 No 103 Lr																																													

FIGURE 10.12

The periodic table of elements, showing atomic numbers and chemical symbols. Each horizontal row is called a period and each vertical column, a group. Our labeling of the groups is one of several schemes in common use. To save space, the inner transition elements have been detached from the main table at the positions indicated by the asterisk and dagger. Elements shown in color are “artificial” in the sense that they do not occur in appreciable amounts on earth.



326 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

FIGURE 10.13

The periodic table showing the levels that are filling in different parts of each period. The filling order for the *d* and *f* subshells is a little irregular; for details, see the periodic table inside the back cover of this book.

1s			1s
2s			2p
3s			3p
4s		3d	4p
5s		4d	5p
6s	*	5d	6p
7s	†	6d	

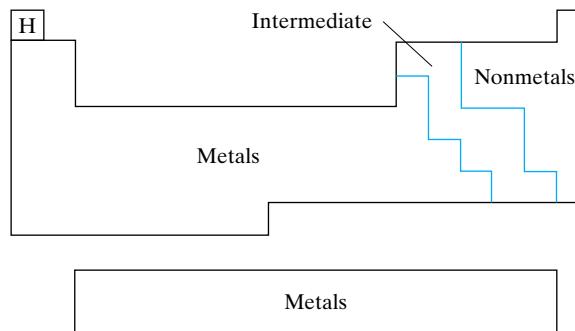
*	5d	4f
†	6d	5f

filled *p* level.* At the extreme right is group VIII (sometimes called group 0), which contains the closed-shell noble gases. Ten interior columns contain the **transition elements**, in which a *d* level is being filled. Since the first *d* level (*3d*) is in the fourth shell, the transition elements begin only in the fourth period. The sixth and seventh shells contain *f* levels (*4f* and *5f*); the **inner transition elements**, in which these *f* levels are being filled, are the elements which are placed in two separate rows at the bottom of the table. The *4f* elements are called the **lanthanides** or **rare earths**, and the *5f* elements are known as **actinides**. Each period of the periodic table can be divided into sections according to the levels being filled. These divisions are shown in Fig. 10.13.

Because of the way the periodic table is constructed, one often finds that elements which are close together in the table have many properties in common. An example is the division of the elements into metals and nonmetals as shown in Fig. 10.14. The elements in the upper right part of the table are all nonmetals, including oxygen, nitrogen, and all the noble gases. Slightly to the left of the nonmetals are a few intermediate elements that are hard to classify; these include the semiconductors germanium and silicon, of such importance in modern electronics. Finally, all elements on the left of the table are metals (except hydrogen, which is generally considered to form a class by itself). We will discuss the properties of metals further in Chapter 13, but we can already understand why they are all found on the left of the periodic table: The high electrical conductivity of metals requires that there be one or more electrons that are easily detached from their atoms. From our knowledge of atomic ionization energies (Fig. 10.11) we would, therefore, expect to find the metals

FIGURE 10.14

The division of elements into metals and nonmetals. Since the properties of the elements vary more or less continuously, the boundaries are only approximate and are sometimes placed a little differently. The elements labeled “intermediate” include the so-called semiconductors — like silicon and germanium — and various elements that are hard to classify — like carbon — which can be a conductor (graphite) or an insulator (diamond).



*The one exception is the noble gas He, with its closed 1s level. You should also be aware that there are several different schemes used for labeling the groups.

toward the beginning and not at the end of each period. In other words, metals should be on the left of the periodic table, exactly as is observed to be the case.

The periodic table is a convenient way to tabulate many different atomic properties. A large table (turned on its side to fit the page better) can be found inside the back cover and includes the atomic number, atomic mass, name, symbol, and electron configuration of every element. With a little use, you will quickly learn to locate elements in the table. If you have any difficulty, you can first find the atomic number, Z , in one of the alphabetical lists of elements in Appendix C and then locate the element itself in the periodic table.

10.9 Excited States of Atoms*

*This section could be omitted without significant loss of continuity.

So far in this chapter, we have discussed only the ground states of multielectron atoms. In this section we give a brief introduction to the excited states. Recall that in the atomic ground state the Z electrons occupy the lowest possible levels consistent with the Pauli principle. To obtain the excited states, we must simply raise one or more of the electrons to higher levels, as illustrated in Fig. 10.15. This picture shows the energy levels of any one electron in the sodium atom ($Z = 11$). The 11 dots show the 11 electrons as they are distributed in the atom's ground state. We obtain the excited states by moving electrons to higher levels; in the picture we have indicated five possibilities — for example, the arrow labeled 1 represents an excitation in which the outer, valence, electron is raised from the $3s$ to the $3p$ level.

Even for an atom as simple as sodium, there is an enormous number of different excited states, corresponding to the many possible arrangements of the Z electrons among the numerous different levels. Fortunately, not all of them are equally important: States in which several electrons have been raised to higher levels are usually less important than those involving a single electron. For example, when an atom absorbs light, the most probable excitations involve moving just a single electron from its ground-state position (as with any one of the arrows in Fig. 10.15). Here we will confine attention to these one-electron excitations.

It is clear in Fig. 10.15 that the excitations that require the least energy are those that involve moving a valence electron to a higher level (arrows 1 or 2,

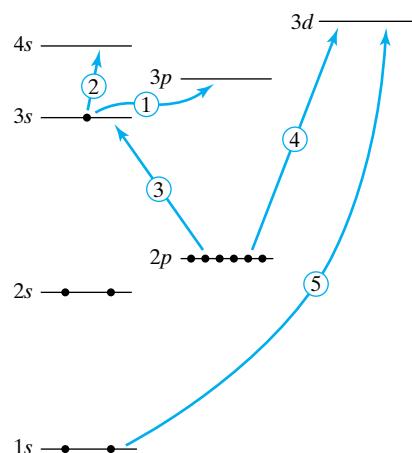
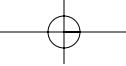


FIGURE 10.15

The energy levels for any one electron in sodium. In the ground state of the atom, the 11 electrons are distributed as shown by the dots. The five arrows indicate five ways in which electrons could be excited to higher levels. (Although the ordering of the levels shown here is correct, they are not drawn to scale; in particular, the $n = 1$ and $n = 2$ levels are much deeper than shown.)



328 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

for example). Since the energy required to excite a valence electron is almost always a few eV, which is the energy of a visible (or nearly visible) photon, these excitations are often called **optical excitations**. Because these are the lowest excitations, a photon (or other projectile) with just enough energy to produce optical excitations cannot produce any other kinds of excitation. Thus it is a simplifying feature of the optical excitations that they often occur in a context where no other kinds of excitation are possible.

If we increase the energy of the photons (or other projectiles) colliding with an atom, they may be able to excite some of the inner electrons (arrows 3, 4, and 5 in Fig. 10.15). It is clear from Fig. 10.15 that the highest-energy excitations are those that involve the innermost, $1s$, electrons. The most striking thing about these is the very large energies involved. We saw in (10.9) that the energy of the $1s$ level is

$$E_{1s} \approx -Z^2 E_R$$

For sodium ($Z = 11$) this is more than 1 keV, and for a heavy atom such as uranium ($Z = 92$) it is more than 100 keV. Because of the Pauli principle, it is impossible to excite a $1s$ electron to any level that is already full. (Thus, in Fig. 10.15 one cannot excite a $1s$ electron to either of the $n = 2$ levels.) Therefore a $1s$ electron can be excited only by lifting it all the way to the valence level or higher. Since the binding energy of these upper levels is only a few eV, the energy required to excite one of the innermost electrons (in all but the very lightest atoms) is several keV or more, the energy of X-ray photons. For this reason, excitations of the inner electrons are called **X-ray excitations**.

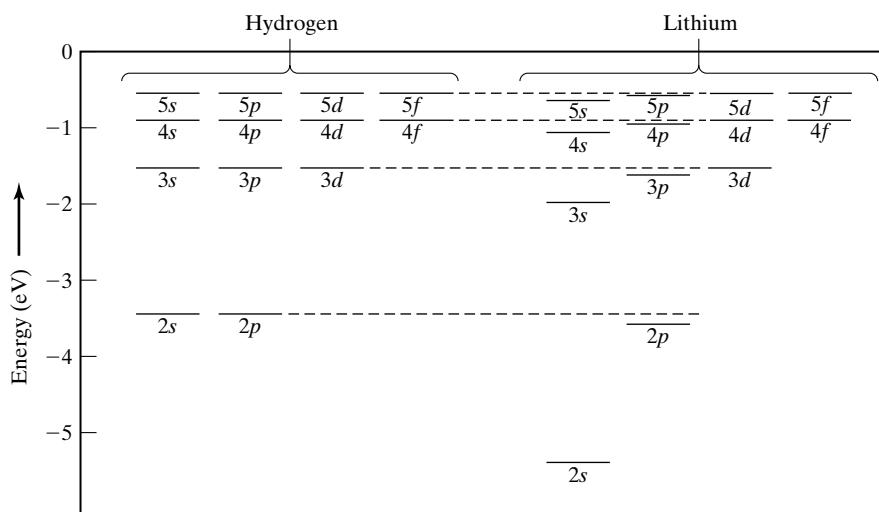
Once a vacancy has been created in an inner shell, any one of the higher electrons can drop down into the vacancy, emitting an X-ray photon as it does so. This is the origin of the X-rays studied by Henry Moseley and described in Section 5.9. The discussion there was based on the Bohr model but depended only on the formula $E_n \approx -Z^2 E_R/n^2$ for the inner energy levels.* Therefore, the results of Section 5.9 carry straight over to a modern quantum treatment and need not be repeated here.

Optical Excitations in Alkali Atoms

To conclude this section, we take a slightly closer look at those excitations involving only the valence electrons — the so-called optical excitations. To simplify the discussion further, we consider only atoms with a single valence electron. These are the closed-shell-plus-one atoms, Li, Na, K, Rb, Cs, Fr, which appear in group I of the periodic table and are called the alkali atoms (or alkali metals, since they all form metals).

In each of the alkali atoms the single valence electron is concentrated outside a core consisting of the nucleus (with charge Ze) and $Z - 1$ electrons, all arranged in spherically symmetric closed shells. The total charge of this core is $+e$, the same as that of the single proton in hydrogen. If the valence electron were entirely outside this core, it would have exactly the same allowed energies as the electron in hydrogen. In reality, the wave function of the valence electron penetrates the core, where the positive charge of the nucleus is less

*Notice that in the discussion of Section 5.9, we took for granted that an electron in one of the levels $n = 2, 3, \dots$ could only drop to the $n = 1$ level after one of the $n = 1$ electrons had been ejected. Thus, we were implicitly assuming the Pauli principle in that discussion.

**FIGURE 10.16**

The energy levels of the H atom compared with the levels of the valence electron in the lithium atom.

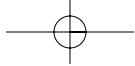
shielded and attracts the electron more strongly. Thus all the levels of the valence electron of an alkali atom lie a little lower than the corresponding levels in hydrogen. Since the $l = 0$ states penetrate closest to the nucleus, they are affected the most. For the states of higher l , the wave functions penetrate the core very little and the energies are very close to the corresponding energies in hydrogen (as was illustrated in Example 10.1).

The close relationship between the optical levels of an alkali atom and the levels of hydrogen is illustrated in Fig. 10.16, which shows the levels of the valence electron in lithium ($Z = 3$). In the ground state the valence electron occupies the $2s$ level, which is appreciably lower (by nearly 2 eV) than the $n = 2$ level in hydrogen, as we would expect. The higher s levels shown are all visibly lower than the corresponding levels of hydrogen, although the difference is smaller for the states of higher n , since these are concentrated at larger radii and penetrate the core less. The difference between corresponding p states is much smaller, and for the d and f states the difference cannot be seen on the scale of Fig. 10.16.

The optical levels of the other alkali atoms have the same general behavior as those of lithium, although the lowest level of the valence electron is one shell higher in each successive atom, Na, K, and so on. Thus the lowest level of the valence electron in sodium ($Z = 11$) is the $3s$ level. The next level in Na is the $3p$ level, about 2.1 eV above the ground state. As we see in the following example, the $3p \rightarrow 3s$ transition in Na produces the yellow light characteristic of the sodium-vapor lamps used to light many streets.

Example 10.2

- (a) The first excited state of the sodium atom is the $3p$ level, 2.10 eV above the $3s$ level. What is the wavelength of the light emitted in the $3p \rightarrow 3s$ transition? (b) Because of the spin-orbit interaction (Section 9.7) the $3p$ level is actually two levels, 2.1×10^{-3} eV apart (whereas the $3s$ is still just a single level). What is the difference $\Delta\lambda$ between the two wavelengths produced by $3p \rightarrow 3s$ transitions in Na?



330 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

(a) The wavelength of a photon with energy $E_\gamma = 2.10 \text{ eV}$ is

$$\begin{aligned}\lambda &= \frac{hc}{E_\gamma} \\ &= \frac{1240 \text{ eV} \cdot \text{nm}}{2.10 \text{ eV}} = 590 \text{ nm}\end{aligned}\quad (10.20)$$

Light with this wavelength is yellow, which is why sodium lamps produce their characteristic yellow glow.

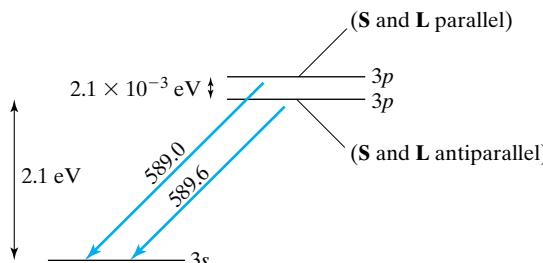
(b) Differentiating (10.20) with respect to E_γ , we find that

$$\begin{aligned}\Delta\lambda &\approx \frac{d\lambda}{dE_\gamma} \Delta E_\gamma = -\frac{hc}{E_\gamma^2} \Delta E_\gamma = -\lambda \frac{\Delta E_\gamma}{E_\gamma} \\ &= -(590 \text{ nm}) \times \frac{2.1 \times 10^{-3}}{2.1} \approx -0.6 \text{ nm}\end{aligned}$$

The actual wavelengths of the two transitions are indicated in Fig. 10.17. It is because of this small “fine structure” splitting that the yellow line of sodium is actually a doublet of two lines, making the sodium spectrum very easy to identify.

FIGURE 10.17

The spin-orbit interaction splits the $3p$ level in Na into two levels, with the states in which S and L are parallel, having slightly higher energy. The wavelengths of the two transitions are shown in nanometers.



CHECKLIST FOR CHAPTER 10

CONCEPT

Independent-particle approx. (IPA)

DETAILS

Each electron moves in the field of the nucleus and the average distribution of the other $Z - 1$ electrons (Sec. 10.2)

IPA potential energy

$U(r) = -Z_{\text{eff}}(r)ke^2/r$ (10.6), where $Z_{\text{eff}}(r) \rightarrow Z$, as $r \rightarrow 0$ and $Z_{\text{eff}}(r) \rightarrow 1$, as $r \rightarrow \infty$

IPA energy levels

See Fig. 10.3

Pauli exclusion principle

No two electrons can occupy the same quantum state (Sec. 10.4)

Bosons and fermions*

Bosons — symmetric multiparticle wave functions (10.15) — do not obey the Pauli principle

Fermions — antisymmetric multiparticle wave functions (10.16) — obey the Pauli principle

Electron configuration

List of occupied energy levels, with numbers of electrons shown as superscripts (Sec. 10.7)

Periodic behavior

Ionization energy, atomic radius, electron affinity
(Fig. 10.11)

Shells

Spatial shell = group of levels with similar radii

Periodic table

Energy shell = group of levels with similar energies

Excited states of atoms *

Closed-shell atom = noble gas (He, Ne, ...)

Closed-shell-plus-one = alkali metal (Li, Na, ...)

Closed-shell-minus-one = halogen (F, Cl, ...)

Table of the elements highlighting periodic properties
(Fig. 10.12)

Optical excitations — valence electrons

X-ray excitations — inner electrons (Sec. 10.9)

PROBLEMS FOR CHAPTER 10

SECTION 10.2 (The Independent-Particle Approximation)

- 10.1** •• Find the electric field \mathcal{E} at $r = a_B$ in the $1s$ state of a hydrogen atom. Compare with the breakdown field of dry air, about 3×10^6 V/m. [Hint: Use Gauss's law; treat the atomic electron as a static charge distribution with charge density $\rho(r) = -e|\psi(r)|^2$; and use the result of Problem 8.43.]
- 10.2** ••• The IPA potential-energy function $U(\mathbf{r})$ is the potential energy "felt" by an atomic electron in the average field of the other $Z - 1$ electrons plus the nucleus. If one knew the average charge distribution $\rho(\mathbf{r})$ of the $Z - 1$ other electrons, it would be a fairly simple matter to find $U(\mathbf{r})$. The calculation of an accurate distribution $\rho(\mathbf{r})$ is very hard, but it is easy to make a fairly realistic guess. For example, one might guess that $\rho(\mathbf{r})$ is spherically symmetric and given by

$$\rho(r) = \rho_0 e^{-r/R}$$

where R is some sort of mean atomic radius. **(a)** Given that $\rho(r)$ is the average charge distribution of $Z - 1$ electrons, find ρ_0 in terms of Z , e , and R . **(b)** Use Gauss's law to find the electric field \mathcal{E} at a point r due to the nucleus and the charge distribution ρ . **(c)** Verify that as $r \rightarrow 0$ and $r \rightarrow \infty$, \mathcal{E} behaves as required by (10.2) and (10.3). [Hint: The integrals needed in parts (a) and (b) are in Appendix B.]

SECTION 10.3 (The IPA Energy Levels)

- 10.3** • **(a)** Estimate the energy of the innermost electron of lead. **(b)** What is its most probable radius? (Appendix C has a list of atomic numbers.)
- 10.4** • Answer the same questions as in Problem 10.3, but for silver.
- 10.5** •• The ground state of sodium ($Z = 11$) has two electrons in the $1s$ level, two in the $2s$, six in the $2p$, and one in the $3s$. Consider an excited state in which the outermost electron has been raised to a $3d$ state (but all of the inner electrons are unchanged). Because

the $3d$ wave function is not very penetrating, you can treat the outer electron as if it were completely outside all the other electrons. **(a)** In this approximation what is the potential-energy function $U(r)$ felt by the outer electron? **(b)** In the same approximation, what should be the energy of an electron in a $3d$ state? Compare your answer with the observed value of -1.52 eV. Why is the observed value lower than your estimate?

- 10.6** •• The ground state of lithium ($Z = 3$) has two electrons in the $1s$ level and one in the $2s$. Consider an excited state in which the outermost electron has been raised to the $3p$ level. Since the $3p$ wave functions are not very penetrating, you can estimate the energy of this electron by assuming it is completely outside both the other electrons. **(a)** In this approximation what is the potential-energy function felt by the outermost electron? **(b)** In the same approximation write the formula for the energy of the outer electron if its principal quantum number is n . **(c)** Estimate the energy of the $3p$ electron in this way and compare with the observed value of -1.556 eV. **(d)** Repeat for the case that the outer electron is in the $3d$ level, whose observed energy is -1.513 eV. **(e)** Explain why the agreement is better for the $3d$ level than for the $3p$. Why is the observed energy for $3p$ lower than that for $3d$?

SECTION 10.4 (The Pauli Exclusion Principle)

- 10.7** • **(a)** How many electrons can be accommodated in an electron energy level with $l = 2$? **(b)** How many if $l = 3$? **(c)** Give a formula (in terms of l) for the number of electrons that can be accommodated in a level with arbitrary l .
- 10.8** • **(a)** Imagine an electron ($spin s = \frac{1}{2}$) confined in a one-dimensional rigid box. What are the degeneracies of its energy levels? **(b)** Make a sketch of the lowest few levels, showing their occupancy for the lowest state of six electrons confined in the same box. (Ignore the Coulomb repulsion among the electrons.)

332 Chapter 10 • Multielectron Atoms; the Pauli Principle and Periodic Table

10.9 •• (a) If the electron had spin $s = \frac{3}{2}$ (but was unchanged in every other respect), how many different orientations would its spin have? **(b)** Sketch energy-level diagrams similar to Fig. 10.6(b) showing how the levels would be occupied in the ground states of helium ($Z = 2$) and lithium ($Z = 3$) if the electron had $s = \frac{3}{2}$.

10.10 •• Repeat Problem 10.8, assuming that the electron had spin $s = \frac{3}{2}$.

10.11 •• Imagine several identical spin-half particles all confined inside the two-dimensional rigid square box discussed in Section 8.3. Assume that the particles do not interact with one another and hence, that the allowed energies for each particle are exactly as shown in Fig. 8.2. **(a)** What are the lowest four allowed energies for any one particle? How many particles can be accommodated in each of these levels, given that they obey the Pauli exclusion principle? [Hint: In figuring degeneracies, do not forget that each particle has two possible orientations of its spin.] **(b)** Assuming that there are six particles in the box, draw an energy-level diagram similar to Fig. 10.6(b), showing the distribution of particles that gives the state of lowest energy for the system as a whole. **(c)** Do the same for the case where there are ten particles in the box.

SECTION 10.5 (Fermions and Bosons*)

10.12 •• The wave function for two spinless particles would have the form $\psi = \psi(x_1, x_2)$. **(a)** Give an example of such a function that is symmetric under particle exchange and normalizable (integral of $|\psi(x_1, x_2)|^2$ over all x_1 and all x_2 is finite). **(b)** Give an example that is antisymmetric. **(c)** Give an example that is neither symmetric nor antisymmetric.

10.13 •• (a) Consider the helium atom to be a fixed point nucleus (charge $2e$) with two spin-half fermion electrons. What is the degeneracy of its ground state? (That is, how many independent states of the whole atom have the ground state energy?) **(b)** Suppose instead that the electron was a spin-half boson. (It is an experimental fact that all spin-half particles are fermions, but there is nothing to stop us imagining a spin-half boson.) What then would be the degeneracy of the He atom's ground state? **(c)** What would the degeneracy be if the two electrons were somehow distinguishable? (For example, they might have slightly different masses.) The different numbers that you should get in the three parts of this problem are examples of the different "statistics" of fermions, bosons, and distinguishable particles. (See Chapter 13.)

SECTION 10.6 (Ground States of the First Few Elements)

10.14 • Draw two energy-level diagrams similar to Fig. 10.7, showing the ground state and first excited state of a boron atom ($_5\text{B}$).

10.15 • Draw four energy-level diagrams, similar to those in Fig. 10.7, to illustrate the ground states of the following atoms: $_5\text{B}$, ${}_9\text{F}$, ${}_{10}\text{Ne}$, ${}_{11}\text{Na}$.

10.16 • Draw two energy-level diagrams similar to Fig. 10.7, showing the ground state and first excited state of a neon atom.

10.17 • Consider the graph of ionization energy against atomic number Z in Fig. 10.8. It is clear that within each shell ($Z = 1$ to 2 and $Z = 3$ to 10), the ionization energy tends to increase with Z . However, there is a small *drop* as one moves from ${}_4\text{Be}$ to ${}_5\text{B}$. Explain this drop. (The second small drop, between ${}_7\text{N}$ and ${}_8\text{O}$, is explained in Section 10.7.)

10.18 •• The first ionization energy of an atom is the minimum energy needed to remove one electron. For helium, this is 24.6 eV. The second ionization energy is the additional energy required to remove a second electron. **(a)** Calculate the second ionization energy of helium. **(b)** What is the total binding energy of helium (that is, the energy to take it apart into two separated electrons and a nucleus)?

10.19 •• In this question you will estimate the total energy of a helium atom. **(a)** What would be the total energy of a helium atom (in its ground state) in the approximation where you ignore completely the electrostatic force between the two electrons? [Hint: In this approximation you can treat each electron separately as if it were in a hydrogen-like ion. The total energy is just the sum of the two separate energies.] **(b)** Your answer in part (a) should be negative (indicating that the system is bound) and *too* negative since you ignored the positive potential energy due to the repulsion between the two electrons. To get a rough estimate of this additional potential energy, imagine the electrons to be in the first Bohr orbit, with radius $a_B/2$ (the appropriate radius for a hydrogen-like ion with $Z = 2$). To minimize their energy, the two electrons would move around the same circular orbit, always on opposite sides of the nucleus, a distance a_B apart. Use this semiclassical model to estimate the potential energy of the two electrons. Combine this with your answer to part (a) to estimate the total energy of the He atom. Compare with the observed value of -79.0 eV.

10.20 •• (a) If the electron had spin $s = \frac{3}{2}$ (but was unchanged in every other respect), how many different orientations would its spin have? **(b)** Sketch energy-level diagrams, showing how the levels would be occupied in the ground states of ${}_4\text{Be}$ and ${}_6\text{C}$ if the electron had $s = \frac{3}{2}$.

SECTION 10.7 (The Remaining Elements)

10.21 • Use the energy-level diagram of Fig. 10.9 to write down the electron configurations of ${}_{30}\text{Zn}$, ${}_{35}\text{Br}$, ${}_{54}\text{Xe}$, ${}_{85}\text{At}$, ${}_{87}\text{Fr}$.

10.22 • Use the energy-level diagram of Fig. 10.9 to find the electron configurations of ${}_{20}\text{Ca}$, ${}_{23}\text{V}$, ${}_{32}\text{Ge}$, ${}_{53}\text{I}$, ${}_{88}\text{Ra}$.

10.23 • Explain the abrupt drops in the ionization energies shown in Fig. 10.11 between Cd and In and between Hg and Tl.

10.24 • Write down the electron configuration for each of the six alkali-metal atoms.

10.25 •• (a) Use the information in Fig. 10.9 to find the ground-state configurations of the following atoms: $_{30}\text{Zn}$, $_{80}\text{Hg}$, $_{37}\text{Rb}$, $_{55}\text{Cs}$. **(b)** What is the total angular momentum of each of these atoms?

10.26 •• (a) Use the information in Fig. 10.9 to find the ground-state configurations of the following atoms: $_{30}\text{Zn}$, $_{31}\text{Ga}$, $_{39}\text{Y}$, $_{48}\text{Cd}$, $_{53}\text{I}$, $_{54}\text{Xe}$, $_{55}\text{Cs}$. **(b)** What is the total spin angular momentum for each of these atoms? **(c)** What are their total orbital angular momenta?

10.27 •• Give the ground-state configurations that the atoms of $_{8}\text{O}$, $_{10}\text{Ne}$, and $_{21}\text{Sc}$ would have if the electron had spin $s = \frac{3}{2}$. (See Problem 10.20.)

10.28 •• The degeneracies of the levels in hydrogen are 2, 8, 18, 32, ... The numbers of electrons in successive shells of multielectron atoms are 2, 8, 8, 18, 18, 32, ... Given the groupings of levels into shells shown in Fig. 10.9, explain the similarities and differences between these two sets of numbers.

SECTION 10.8 (The Periodic Table)

10.29 • Use the periodic table inside the back cover to write down the full electron configurations of $_{3}\text{Li}$, $_{10}\text{Ne}$, $_{12}\text{Mg}$, $_{19}\text{K}$, $_{28}\text{Ni}$, and $_{48}\text{Cd}$. [Hint: That periodic table gives the configuration of the outer-shell electrons only; to find the full configuration, you must add the configuration of the preceding closed-shell element. For example, for $_{12}\text{Mg}$, add the configuration of $_{10}\text{Ne}$.]

10.30 • Use the periodic table inside the back cover to find the full configurations of the following atoms: iron, silver, iodine, polonium. (See the hint for Problem 10.29, and use the alphabetical list in Appendix C if necessary.)

10.31 • Use the periodic table inside the back cover to find the names, atomic numbers, and full electron configurations of the following atoms: Ga, Xe, W, At, Md, Sg. (See the hint for Problem 10.30.)

10.32 • (a) Find the ground-state configurations of nickel and copper from the periodic table inside the back cover. **(b)** Draw energy-level diagrams, similar to those in Fig. 10.7, to illustrate these two ground states. (Note: In Fig. 10.9 we showed the $4s$ level below $3d$ since it fills first. Nevertheless, $4s$ and $3d$ are very close together; as Z increases, the $3d$ level becomes almost exactly degenerate with $4s$, and $4s$ can lose one electron to $3d$ — as happens in copper.)

10.33 •• Because of the way that atomic properties vary smoothly along the rows and columns of the periodic table, one can often predict the properties of an element from the known properties of its neighbors. (This is how Mendeleev predicted the existence and several properties of the elements now called scandium, gallium, and germanium.) **(a)** The ionization energies of $_{20}\text{Ca}$ and $_{38}\text{Sr}$ are 6.11 eV and 5.70 eV. If one

guessed that ionization energies should change linearly as one moves through a group, what would one predict for the ionization energy of $_{56}\text{Ba}$? (The observed value is 5.21 eV.) **(b)** Use the same argument to predict the electron affinity of $_{35}\text{Br}$, given that $_{17}\text{Cl}$ has electron affinity 3.61 eV and $_{53}\text{I}$ has 3.06 eV. (The observed value is 3.36 eV.) **(c)** Predict the radius of the $_{9}\text{F}$ atom, given that $_{7}\text{N}$ and $_{8}\text{O}$ have radii of 0.075 and 0.065 nm. (The actual value for $_{9}\text{F}$ is 0.057 nm.)

10.34 •• (a) The electron affinities of $_{26}\text{Fe}$ and $_{27}\text{Co}$ are 0.163 and 0.661 eV. Using arguments similar to those outlined in Problem 10.33, predict the electron affinity of $_{28}\text{Ni}$. (The observed value is 1.156 eV.) **(b)** In the same way, predict the boiling point of $_{32}\text{Ge}$, given that the boiling point of $_{14}\text{Si}$ is 3540 K and that of $_{50}\text{Sn}$ is 2876 K. (The observed value is 3107 K.) **(c)** Similarly, predict the melting point of $_{86}\text{Rn}$, given that those of $_{36}\text{Kr}$ and $_{54}\text{Xe}$ are 116 K and 161 K. (The observed value is 202 K.)

10.35 •• The simple estimation of atomic properties as described in Problems 10.33 and 10.34 does not always work very well. Here is a well-known example where it is fairly unsuccessful: The electron affinities of $_{17}\text{Cl}$ and $_{35}\text{Br}$ are 3.62 and 3.36 eV. Assuming that electron affinity varies linearly within the groups of the periodic table, what would you predict for the electron affinity of $_{9}\text{F}$? The observed value is 3.40 eV.

SECTION 10.9 (Excited States of Atoms*)

10.36 • The ground-state configuration of the lithium atom is $1s^2 2s^1$. Give the configurations of the lowest five excited levels. (See Fig. 10.16, and ignore fine structure.)

10.37 • Give the configurations of the lowest four levels of a sodium atom. (See Fig. 10.15, and ignore fine structure.)

10.38 • The spectrum of atomic lithium has a red line at $\lambda = 671$ nm, arising from the transition $2p \rightarrow 2s$. On close inspection, this line is seen to be a doublet of lines separated by 0.0152 nm. What is the fine-structure splitting of the $2p$ energy level in lithium?

10.39 • Give the configurations of the lowest six levels of the He atom.

COMPUTER PROBLEM

10.40 ••• If you have not already done it, do Problem 10.2. What is the potential energy $U(r)$ of an electron in this electric field? Use any graphing software to make a plot that shows $U(r)$ as well as the expected forms (10.4) and (10.5) for the case that $Z = 4$. Does your plot confirm the behavior sketched in Fig. 10.1?

Chapter 11

Atomic Transitions and Radiation

- 11.1** Introduction
- 11.2** Radiation by Classical Charges
- 11.3** Stationary States and Transitions
- 11.4** More Quantum Formalism*
- 11.5** Transitions; Time-Dependent Perturbation Theory*
- 11.6** A Brief Review
- 11.7** Spontaneous Emission
- 11.8** Atomic Selection Rules
- 11.9** Lasers
- 11.10** Further Properties of Lasers*

Problems for Chapter 11

*These sections could be omitted without significant loss of continuity.

11.1 Introduction

In this chapter we discuss the emission and absorption of radiation by atoms making transitions between different energy levels. Our general observations will also apply to a wide variety of other quantum systems, including ions, molecules, and nuclei. We begin with a classical description of radiation and apply it to a classical model of the atom with electrons in circular orbits. In doing so, we discover some useful concepts and see how the classical view of an atom fails to agree with observation. Next we will see how radiation striking a quantum atom can cause transitions between quantum states. This process accounts for the emission or absorption of a photon by an atom exposed to external radiation, and also for the spontaneous emission of a photon by an isolated atom in an excited state. Armed with this understanding, we conclude the chapter by describing one of the most exciting applications of quantum theory, the laser.

11.2 Radiation by Classical Charges

An accelerating charge produces electric and magnetic fields that change with time. Faraday's law of induction tells us that a changing magnetic field induces an electric field, and, as first proposed by Maxwell, a changing electric field induces a magnetic field. This suggests that an accelerating charge can bring about an intertwined process in which changing electric and magnetic fields continually induce one another, with the fluctuations in these fields moving

Section 11.2 • Radiation by Classical Charges **335**

away from the charge as electromagnetic radiation. This suggestion proves to be correct: All of the many forms of electromagnetic radiation around us — light from a lightbulb or the sun, radio waves from our favorite radio station, X-rays in our doctor's office — are produced by accelerating charges.

In Maxwell's theory of electromagnetic radiation these ideas are made precise by showing that the laws of magnetic and electric induction can be combined to give a wave equation for both the magnetic and electric fields. The wave speed c for disturbances of either field is given by the combination of electromagnetic constants

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3 \times 10^8 \text{ m/s}$$

which equals the observed speed of light. This suggested to Maxwell, what has since been amply verified, that light consists of electromagnetic waves in which electric and magnetic fields travel together.

To get some picture of how an accelerating charge produces radiation, let us examine the electric field of a single charge that is abruptly put into motion. In Fig. 11.1(a) we show the initially static charge and its radial field. When the charge is put into motion, the news of its changed position moves outward at the speed of light, so distant portions of the field lines are offset from near portions. The transition zone between the near and far fields necessarily contains a transverse component, as shown in Fig. 11.1(b) and (c).* While the radial component of the electric field falls like $1/r^2$ it can be shown that the transverse component falls like $1/r$. Consequently, at large distances it is the transverse component that dominates and carries radiated energy away from the charge.

The total power P radiated by any single charge q (moving nonrelativistically†) can be shown to be

$$P = \frac{2kq^2a^2}{3c^3} \quad (11.1)$$

where a is the charge's acceleration. This formula accurately describes the power radiated by any macroscopic system of moving charges. For example, in TV or radio broadcasting, electric charges are made to oscillate inside the rods of an antenna, and the resulting radiated power is given by (11.1) (see Problem 11.2). Notice that the power (11.1) depends on the acceleration a . Thus a charge moving at constant velocity does not radiate.‡ We should also mention that with an assembly of *many* accelerating charges, the fields produced by the different charges can sometimes interfere destructively, with no net radiated power. For example, consider a uniform ring of charge rotating at a constant rate; this amounts to a steady current loop and does not radiate any power.

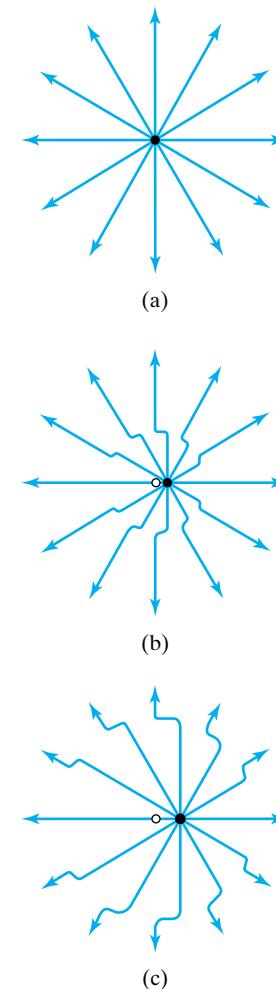


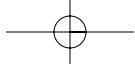
FIGURE 11.1

(a) Electric field lines from a static charge are radial. (b) When the charge is given an abrupt kick to the right, changes in its electric field propagate outward at speed c ; distant portions of the field still point outward from the original (open circle) position. (c) The transverse disturbance linking near and far fields continues to move radially outward as the charge coasts forward.

*Fig. 11.1 shows the field lines for a charge with speed much less than c . The pattern for relativistic speeds is more complicated. For a beautiful discussion of the fields from accelerated charges, see Edward Purcell, *Electricity and Magnetism* (McGraw-Hill, 1964), Chapter 5.

† To see how (11.1) is modified when the motion is relativistic, see Problem 11.7.

‡ To see why no radiation occurs when \mathbf{v} is constant, consider the inertial frame in which $\mathbf{v} = 0$. In this frame the magnetic field is zero and the electric field is constant, so no radiation occurs. It follows from the principle of relativity that no radiation occurs in any frame.



336 Chapter 11 • Atomic Transitions and Radiation

The formula (11.1) and the classical theory from which it derives can sometimes be applied to microscopic systems as well. For example, when an electron in an X-ray tube collides with the anode, it undergoes rapid deceleration. The classical theory correctly predicts that this deceleration will produce the X-ray radiation, known as bremsstrahlung, or braking radiation (Section 4.4). Since the classical theory takes no account of the quantization of radiation, it cannot, of course, be correct in every detail. For example, classical theory predicts that some bremsstrahlung X-rays will be produced at all frequencies. In reality, each photon carries energy hf , and no photons can be produced for which hf exceeds the kinetic energy K of the incident electron. Therefore, the spectrum of emitted X-rays is cut off abruptly at $f = K/h$ as predicted by the Duane–Hunt law of Section 4.5.

When the formula (11.1) was applied to a classical atom, difficulties arose almost immediately, and it was this problem that suggested to Bohr that classical physics needed modification, as described in Section 5.4. The problem in question is that an electron orbiting in a classical atom is necessarily accelerating. Therefore, according to (11.1), it must be radiating electromagnetic energy. In the following example we see that the rate of radiation predicted by (11.1) is so large that a classical atom would collapse completely in a time on the order of 10^{-11} seconds.

Example 11.1

Find the power that would be radiated by a classical electron in the $n = 1$ Bohr orbit of a hydrogen atom.

The orbiting electron has a centripetal acceleration $a = v^2/r$ and must therefore radiate. Perhaps the simplest way to find the acceleration is to note that it is given by Newton's second law as $a = F/m$, with $F = ke^2/r^2$. Substituting into (11.1), we find that

$$P = \frac{2ke^2}{3c^3} \left(\frac{ke^2}{mr^2} \right)^2 = \frac{2(ke^2)^3 c}{3(mc^2)^2 r^4} \quad (11.2)$$

or, putting $r = a_B = 5.3 \times 10^{-2}$ nm,

$$\begin{aligned} P &= \frac{2 \times (1.44 \text{ eV} \cdot \text{nm})^3 \times (3 \times 10^{17} \text{ nm/s})}{3 \times (5.1 \times 10^5 \text{ eV})^2 \times (5.3 \times 10^{-2} \text{ nm})^4} \\ &= 2.9 \times 10^{11} \text{ eV/s} \end{aligned} \quad (11.3)$$

where we have again used the useful combination $ke^2 = 1.44 \text{ eV} \cdot \text{nm}$. Since typical energies of electrons in atoms are a few eV, the rate (11.3) is an extremely rapid energy loss. (For more details, see Problems 11.6 and 11.15.)

The result (11.3) shows that the orbit of an electron with an energy of a few eV will change radically in a time of order 10^{-11} seconds. More precisely, as the electron loses energy, its orbital radius will shrink, and once r reaches zero, the atom will have collapsed completely. As you can check in Problem 11.15, the time τ for a hydrogen atom in its ground state to collapse in this way would be

$$\tau = 1.6 \times 10^{-11} \text{ s} \quad (11.4)$$

This dramatic instability of the classical atom was what led Bohr to postulate the existence of quantized orbits to which the classical laws of radiation did not apply. As we saw in Chapter 7 (and review in Section 11.3), the Schrödinger theory avoids this problem since it predicts that atoms have *stationary* charge distributions.

11.3 Stationary States and Transitions

In Section 11.2 we saw that the radiation from a classical atom would lead to a rapid collapse of the electrons' orbits. We now turn to a proper quantum treatment of the atomic electrons and will see how this disastrous collapse is avoided. As we discussed in Section 7.3, the wave function $\Psi(\mathbf{r}, t)$ of an electron generally depends on time as well as the spatial coordinates. In particular, suppose that at time $t = 0$ the wave function is a stationary state $\Psi(\mathbf{r}, 0) = \psi_n(\mathbf{r})$ with energy E_n ; that is, ψ_n satisfies the time-independent Schrödinger equation with $E = E_n$. Then at any later time t the complete wave function $\Psi(\mathbf{r}, t)$ is a product of $\psi_n(\mathbf{r})$ and the oscillatory phase factor $e^{-iE_n t/\hbar}$:

$$\Psi(\mathbf{r}, t) = \psi_n(\mathbf{r})e^{-iE_n t/\hbar} \quad (11.5)$$

as in Eq. (7.11). In the problems considered so far, we have been able to restrict attention to the spatial wave function $\psi_n(\mathbf{r})$. However, since transitions are a distinctly time-dependent phenomenon, we must now examine the full time-dependent wave function $\Psi(\mathbf{r}, t)$.

As we discussed in Chapter 6, the probability density P of an electron is given by the square of the absolute value of its wave function

$$P(\mathbf{r}, t) = |\Psi(\mathbf{r}, t)|^2 \quad (11.6)$$

which for the wave function (11.5), becomes

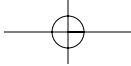
$$P(\mathbf{r}, t) = |\psi_n(\mathbf{r})|^2 \times |e^{-iE_n t/\hbar}|^2 = |\psi_n(\mathbf{r})|^2 \quad (11.7)$$

since $|e^{-i\theta}| = 1$. (See Sec. 7.3.) We see in (11.7) that the probability density (and hence the charge density) of an electron in the state (11.5) does not depend on time. It is for this reason that these states are called *stationary states*. Their charge distributions are stationary and hence do not radiate.*

The simple Schrödinger theory outlined here successfully explains the existence of stable atomic states. Unfortunately, it goes too far, since it predicts that all states of definite energy should be perfectly stable. Thus, our task now is to explain how transitions *can* occur, in which an atom or other quantum system can move from one energy level to another. We will describe the theory of transitions in the next two sections, but let us first sketch the main features here.

If the initial wave function $\Psi(\mathbf{r}, 0) = \psi_n(\mathbf{r})$ really is a solution of the time-independent Schrödinger equation, including the correct potential energy U of the isolated atom, then the full time-dependent wave function is given

*You might object that in a stationary state with $\ell \neq 0$ the charge distribution may be stationary, but the electron is surely circulating since it has nonzero orbital angular momentum. This is correct, but this situation corresponds to a steady current loop and still does not radiate.



338 Chapter 11 • Atomic Transitions and Radiation

by (11.5) and the atom remains locked in the same state for all time. For transitions to occur, there has to be some outside influence that adds an additional term W to the potential energy. For example, we might switch on an electric field \mathcal{E} in the x direction. In this case the electron would acquire an additional potential energy

$$W = e\mathcal{E}x \quad (11.8)$$

This extra potential energy is called a **perturbation**, since it perturbs the system and causes transitions. Perturbations can come about in several ways; for example, a pulse of radiation directed at an atom can produce a brief oscillatory electric field causing a potential energy of the form (11.8). The passage of a second atom close by the atom of interest can change the potential energy of the latter and cause transitions. Finally, as we will argue in Section 11.7, even an atom that is perfectly isolated is subject to tiny electric and magnetic fields called vacuum fluctuations or zero-point fields, and it is these vacuum fluctuations that cause an excited but isolated atom to emit a photon and drop to a lower energy level.

It is easy to understand in a general way how a perturbation like (11.8) can cause transitions. The wave function ψ_n is a solution of the Schrödinger equation, including just the potential energy U of the isolated atom. It is *not* a solution of the Schrödinger equation with the full potential energy $U + W$ and is not, therefore, a stationary state of the actual problem including the perturbation. Therefore a system that starts out in the state ψ_n can change to some other state ψ_m .

The theory of transitions caused by a perturbation is called **time-dependent perturbation theory**. As we will describe in Sec. 11.5, when a system starts out at time $t = 0$ in a state $\Psi(\mathbf{r}, 0) = \psi_n(\mathbf{r})$ and is subjected to a perturbation W , there is a definite probability $P(n \rightarrow m)$ that a subsequent measurement will find the system in a different state $\psi_m(\mathbf{r})$. In some simple situations, we will see how this probability can actually be calculated, but the main purpose of the next two sections is just to see how quantum mechanics can account for the occurrence of transitions. If you feel happy to accept the claims just made, then you can skip these two sections without significant loss of continuity.

11.4 More Quantum Formalism*

*This section can be omitted without significant loss of continuity.

Before we describe the quantum-mechanical theory of transitions, it is useful to introduce some changes of notation and a couple of new ideas. These are developments that you will certainly meet if you take another course in quantum mechanics, and now is a good opportunity to introduce them.

The Hamiltonian Operator

Our first development concerns the way that we write the time-independent Schrödinger equation,

$$\frac{d^2\psi}{dx^2} = \frac{2m}{\hbar^2} [U(x) - E]\psi \quad (11.9)$$

(For simplicity, we'll consider a one-dimensional system for now.) The form (11.9) is convenient for study as a differential equation, but it is not the form in which the equation is normally written in more advanced work, where it is usually reorganized to read

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x) \right] \psi = E\psi \quad (11.10)$$

The quantity in the square brackets on the left is a **differential operator**, or just **operator**: "Operating" on a function ψ , it gives the function $-(\hbar^2/2m)\psi'' + U\psi$. This operator is so important, it is given a name, the **Hamiltonian operator**, or just **Hamiltonian**,* and is denoted H

$$H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x) \quad (11.11)$$

The Hamiltonian is intimately connected to the energy of the system. For example, if ψ is the plane wave $\psi = e^{ikx}$,

$$H\psi = \left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U \right] e^{ikx} = \left[\frac{\hbar^2 k^2}{2m} + U \right] \psi = [K + U]\psi \quad (11.12)$$

In other words, the operator H acting on a plane wave gives $(K + U)$ times the plane wave. More important, the time-independent Schrödinger equation can now be written as

$$H\psi = E\psi \quad (11.13)$$

from which we see that for any wave function ψ that has a definite energy E , the differential operator H acting on ψ gives simply E times ψ . In this sense, the Hamiltonian operator of any system determines the allowed energies of the system.

One of the great advantages of this new notation is this: If we were to consider a three-dimensional particle, the Schrödinger equation in the form corresponding to (11.10) would become

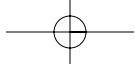
$$\left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + U(x) \right] \psi = E\psi \quad (11.14)$$

But if we agree to define the Hamiltonian for a three-dimensional particle as

$$H = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + U(x)$$

the three-dimensional Schrödinger equation takes the same simple form, $H\psi = E\psi$, as in one dimension. No matter how complicated our system, we

*The origin of the name is this: In classical mechanics, the Hamiltonian function is a function that gives the energy of a system (named after the Irish mathematician William Rowan Hamilton). As we will argue directly, the Hamiltonian operator plays an analogous role in quantum theory.



340 Chapter 11 • Atomic Transitions and Radiation

can always define the Hamiltonian H so that the Schrödinger equation takes this simple form, $H\psi = E\psi$. Of course, the operator H may become very complicated, but the great advantage of our new notation is that we can discuss the Schrödinger equation for *all conceivable systems* at the same time in the one simple form $H\psi = E\psi$. For those readers who haven't seen this kind of equation before, we should emphasize the form of the equation:

$$(\text{differential operator}) \times \text{function} = \text{number} \times (\text{same function})$$

This type of equation is called an **eigenvalue equation**, and the values of E for which it has solutions (namely, the allowed energies) are called **eigenvalues**.

The equation that determines the time development of an arbitrary quantum state (of definite energy or otherwise) is the *time-dependent* Schrödinger equation, which we introduced in Section 7.11.* For a one-dimensional particle, this is

$$i\hbar \frac{\partial}{\partial t} \Psi(x, t) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right] \Psi(x, t) \quad (11.15)$$

which we can immediately rewrite in terms of the Hamiltonian operator as

$$i\hbar \frac{\partial}{\partial t} \Psi = H\Psi \quad (11.16)$$

and in this form, the equation applies to *any* system (although, of course, the details of the operator H vary from system to system). This is the equation we will use to study how atoms make their transitions from one energy level to another.

A Useful Approximation

It is very seldom that one can solve the time-dependent Schrödinger equation exactly, and physicists have developed many schemes for finding approximate solutions. In this chapter we will use just one simple scheme, which lets us find Ψ at a time $t + \Delta t$ in terms of its value at time t , at least if Δt is small. This scheme is based on the familiar approximation

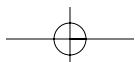
$$f(t + \Delta t) \approx f(t) + \Delta t \frac{df}{dt}$$

which follows from the definition of the derivative. In the case of the wave function Ψ , the time derivative $\partial\Psi/\partial t$ is given by the time-dependent Schrödinger equation (11.16) and the corresponding approximation becomes

$$\Psi(t + \Delta t) \approx \Psi(t) + \frac{\Delta t}{i\hbar} H\Psi(t) \quad (11.17)$$

Notice that Ψ depends on other variables besides t [for a particle in one dimension $\Psi = \Psi(x, t)$ as in (11.15); for a particle in three dimensions, $\Psi = \Psi(x, y, z, t)$; and so on], but for now, it is convenient to omit these.

*If you skipped Section 7.11, you should return to read it now, before continuing with this and the next section.



The approximation (11.17) tells us what gets added to Ψ as the time advances by Δt . It is interesting to check this for the case that Ψ happens to be a stationary state: Suppose that at $t = 0$, $\Psi(0) = \psi_n$, a state with definite energy E_n , satisfying the time-independent Schrödinger equation $H\psi_n = E_n\psi_n$. According to (11.17), the wave function a short time later is

$$\Psi(\Delta t) \approx \Psi(0) + \frac{\Delta t}{i\hbar} H\psi_n = \psi_n - \frac{i\Delta t}{\hbar} E_n \psi_n = \left(1 - \frac{iE_n \Delta t}{\hbar}\right) \psi_n \quad (11.18)$$

That is, the new wave function at time Δt is the original function ψ_n multiplied by a (complex) number and represents the same physical state (as we knew it must, since it is a stationary state). In fact, the multiplicative factor that appears in the last expression of (11.18) is easily understood since we know the exact solution of the time-dependent Schrödinger equation in this case: As in (11.5)

$$\Psi(t) = e^{-iE_n t/\hbar} \psi_n = \left(1 - \frac{iE_n t}{\hbar} - \frac{E_n^2 t^2}{2\hbar^2} + \dots\right) \psi_n \quad (11.19)$$

where the sum in parentheses is just the Taylor series for the exponential factor in the previous expression. Comparing (11.18) with (11.19), we see that the approximation (11.18) is just the first two terms of the Taylor series for the exact form of $\Psi(\Delta t)$ and is certainly a good approximation for small time intervals Δt .

Completeness of the Stationary-State Wave Functions

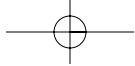
The stationary states of any quantum system are the solutions of the time-independent Schrödinger equation, $H\psi_n = E_n\psi_n$. It is a remarkable fact, though not one we can prove here, that *any* wave function ψ of the system (stationary state or not) can be expanded in terms of these stationary-state wave functions. That is, any wave function ψ can be written as

$$\psi(x) = \sum_n A_n \psi_n(x) \quad (11.20)$$

where the coefficients A_n are constant complex numbers, which depend on the particular function $\psi(x)$ we are expanding, (and we have written the expansion for the case of a one-dimensional particle, though it is equally valid for any quantum system). A set of functions with this property, that they can be used to expand any function, is often called a **complete set of functions**, or a **basis set**. Thus, we can rephrase our claim to say that, for any quantum system, the stationary-state wave functions are a complete set for the expansion of any wave function of the system. The significance of the expansion coefficients A_n is this: $|A_n|^2$ is the probability that a measurement of the energy will result in the value E_n .

Example 11.2

Recall the stationary-state wave functions for a particle of mass m in a one-dimensional rigid box (the infinite square well) of width a , and write down the expansion of an arbitrary wave function $\psi(x)$ in terms of these stationary-state functions. What are the expansion coefficients A_n if $\psi(x)$ is in fact the



342 Chapter 11 • Atomic Transitions and Radiation

ground-state wave function? What if $\psi(x)$ is the first excited state? What if $\psi(x)$ is a 50–50 mixture of the lowest two states?

The stationary-state wave functions for the rigid box are given in (7.60):

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \quad [n = 1, 2, 3, \dots] \quad (11.21)$$

So the expansion (11.20) reads

$$\psi(x) = \sum_{n=1}^{\infty} A_n \sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \quad (11.22)$$

which you may recognize as the Fourier expansion of an arbitrary function in terms of sine functions, as sketched in Chapter 6. This example illustrates a common occurrence: For a particular quantum system (the infinite square well here), the completeness of the stationary-state functions is often derivable from some standard mathematical theorem (Fourier's expansion theorem in this case). In general, however, it is just a postulate of quantum theory that all systems have this property.

If the wave function $\psi(x)$ is known to be the ground state $\psi_1(x)$, then clearly

$$A_1 = 1, \quad \text{and} \quad A_2 = A_3 = \dots = 0$$

This shows that the probability of finding the energy $E = E_1$ is $|A_1|^2 = 1$, while the probability of finding any other value is 0. Similarly, if $\psi(x)$ is known to be the first excited state $\psi_2(x)$, then

$$A_1 = 0, \quad A_2 = 1, \quad \text{and} \quad A_3 = A_4 = \dots = 0$$

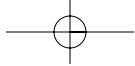
Finally, if $\psi(x)$ is known to be a 50–50 mixture of the two lowest states, the probabilities of finding $E = E_1$ or $E = E_2$ must each be a half, so $|A_1|^2 = |A_2|^2 = 1/2$; that is, both A_1 and A_2 must have absolute value $1/\sqrt{2}$. This does not quite fix the two coefficients, and indeed there are several distinct wave functions that can all be described as 50–50 mixes of the two lowest states. One example would have

$$A_1 = A_2 = 1/\sqrt{2}, \quad \text{and} \quad A_3 = A_4 = \dots = 0$$

but another would have $A_1 = -A_2 = 1/\sqrt{2}$ and so on.

Orthogonality of the Stationary-State Wave Functions

If we know the stationary-state wave functions $\psi_n(x)$ for some system and we are given an arbitrary function $\psi(x)$, then it is surprisingly easy to find the coefficients A_n in the expansion (11.20). This is because of a remarkable property of any set of stationary-state wave functions, called **orthogonality**: If

Section 11.5 • Transitions; Time-Dependent Perturbation Theory **343**

$\psi_n(x)$ (with $n = 1, 2, \dots$) are the stationary states of any quantum system, then it can be shown that*

$$\int_{-\infty}^{\infty} \psi_m(x)^* \psi_n(x) dx = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases} \quad (11.23)$$

That this integral equals 1 if $m = n$ is just a statement that the wave functions are normalized (as we assume throughout this discussion). That it is zero if $m \neq n$ is a new result, often expressed by saying that ψ_m is **orthogonal** to ψ_n , since the integral in (11.23) is in some ways analogous to the scalar product $\mathbf{a} \cdot \mathbf{b}$ of two vectors \mathbf{a} and \mathbf{b} , and the scalar product is zero if the vectors are orthogonal. For a proof of the orthogonality property for one particular system, see Problem 11.17.

Finding the Expansion Coefficients

Armed with the orthogonality relation (11.23), it is a simple matter to find the coefficients A_n in the expansion (11.20) of any given function $\psi(x)$. Multiplying both sides of (11.20) by $\psi_m(x)^*$ and integrating, we find that

$$\int_{-\infty}^{\infty} \psi_m(x)^* \psi(x) dx = \sum_{n=0}^{\infty} A_n \int_{-\infty}^{\infty} \psi_m(x)^* \psi_n(x) dx$$

By (11.23), all of the integrals on the right are zero, except the one with $n = m$, which is equal to 1. Thus the whole infinite series collapses to a single term, equal to A_m , and we conclude that

$$A_m = \int_{-\infty}^{\infty} \psi_m(x)^* \psi(x) dx \quad (11.24)$$

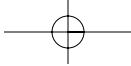
If you did either of Problems 6.35 or 6.36, you will recognize the result (11.24) as the method used to find the coefficients in a Fourier series. We will use the result (11.24) in the next section to find the probabilities of transitions between different energy levels of a quantum system.

11.5 Transitions; Time-Dependent Perturbation Theory*

*Like Sec. 11.4, this section can be omitted without significant loss of continuity. However, you do need to have read 11.4 before reading 11.5.

We are now ready to see how an atom or other quantum system in a stationary state ψ_n can make a transition to another stationary state ψ_m . First, for reasons that will emerge directly, we will call the Hamiltonian operator for

*We state the result here for a particle in one dimension, but it extends in a natural way to any quantum system. For example, for a three-dimensional particle, we have only to replace the integral over x by an integral over x, y , and z . We have stated the orthogonality property for the case that the energy levels are nondegenerate; if there are any degeneracies, the statement needs a little more care, though the practical consequences are the same.



344 Chapter 11 • Atomic Transitions and Radiation

the isolated system H_0 . For example, if our system is an electron in one dimension, then

$$H_0 = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x)$$

If U is the exact potential energy of our isolated system and if the system is initially in the state ψ_n (which satisfies the time-independent Schrödinger equation $H_0\psi_n = E_n\psi_n$), we know that the full wave function for all time has the form (11.5),

$$\Psi(t) = \psi_n e^{-iE_n t/\hbar}$$

Therefore, the system remains in the same stationary state ψ_n for all time, and no transitions can occur. For transitions to occur, the true Hamiltonian H of the system must contain an additional term,

$$H = H_0 + W \quad (11.25)$$

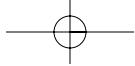
At the end of Sec. 11.3, we mentioned several ways in which this extra perturbation W can come about. For the moment, let us suppose it arises from an electric field \mathcal{E} in the x direction, switched on at time 0, in which case the electron acquires the additional potential energy (11.8)

$$W = e\mathcal{E}x \quad (11.26)$$

The situation is now this: At time $t = 0$, our system is in the state $\Psi(0) = \psi_n$. The state ψ_n is a stationary state, but it is a stationary state of the “wrong” Hamiltonian, H_0 . The true Hamiltonian is $H = H_0 + W$, and to find what happens to the system, we must solve the time-dependent Schrödinger equation using this H . The approximate method of solution that we will describe is called **time-dependent perturbation theory**. To simplify our discussion, we will suppose that the perturbation W is switched on for a short time Δt , in which case we can use the approximate solution (11.17) to find $\Psi(\Delta t)$ after the perturbation is switched off:

$$\begin{aligned} \Psi(\Delta t) &\approx \Psi(0) + \frac{\Delta t}{i\hbar} H \Psi(0) = \psi_n + \frac{\Delta t}{i\hbar} (H_0 + W) \psi_n \\ &= \left(1 - \frac{iE_n \Delta t}{\hbar}\right) \psi_n - \frac{i \Delta t}{\hbar} W \psi_n \end{aligned} \quad (11.27)$$

Comparing with (11.18), you can see that the first term here is just what would have evolved in the absence of any perturbation; in particular, it is proportional to ψ_n , so has not changed the physical state of the system at all. On the other hand, the second term is an additional term, which can represent a small component of a *different* stationary state ψ_m . That is, because of this extra term resulting from the perturbation W that we switched on, there is a nonzero probability that we will find the system in a different state ψ_m . Before we discuss this important result any more, let us illustrate it with a simple example.

Section 11.5 • Transitions; Time-Dependent Perturbation Theory **345****Example 11.3**

An electron is initially in the ground state ψ_1 of a one-dimensional rigid box. At time $t = 0$ we switch on an electric field \mathcal{E} in the x direction, giving the electron an additional potential energy $W = e\mathcal{E}x$. Describe the form of the electron's wave function a short time Δt later when we switch the perturbation off.

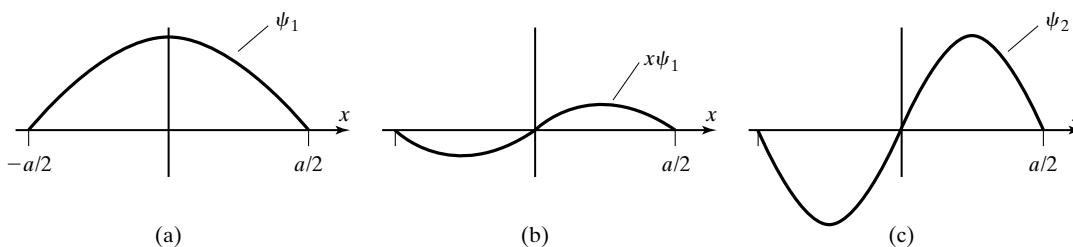
According to (11.27), with W replaced by $e\mathcal{E}x$, the wave function at time t is

$$\Psi(\Delta t) \approx A\psi_1 - \frac{ie\mathcal{E}\Delta t}{\hbar} x\psi_1 \quad (11.28)$$

where A denotes the multiplier in parentheses of (11.27). The first term is just a multiple of the original wave function ψ_1 and is what would have evolved in the absence of any perturbation. The second is the extra term, caused by the perturbing field \mathcal{E} . Its significance is illustrated in Fig. 11.2. Part (a) of that figure shows the original wave function ψ_1 . (It is convenient in this discussion to place the origin at the center of the well, so the edges are now at $\pm a/2$.) Part (b) shows the extra term of (11.28), which is proportional to $x\psi_1$. The amplitude of this function depends on the parameters of the problem, but is certainly small (since our approximate solution requires that Δt be small). The important point is that the extra term shown in Fig. 11.2(b) has almost exactly the shape of the first excited state ψ_2 , which is plotted for comparison in part (c). The main message of Figure 11.2 is that switching on the perturbation $W = e\mathcal{E}x$ has added on to the initial wave function ψ_1 a small term that closely approximates ψ_2 (times some constant), so that the full wave function now has the form

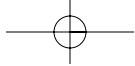
$$\Psi(\Delta t) \approx A\psi_1 + B\psi_2 \quad (11.29)$$

There is now a small probability — namely, $|B|^2$ — that a measurement of the system will find it in the state* ψ_2 .

**FIGURE 11.2**

(a) The ground-state wave function ψ_1 for an electron in an infinite square well of width a , centered on the origin. (b) An electric field \mathcal{E} switched on briefly adds to the wave function of part (a) a term proportional to $x\psi_1$. (c) The extra term shown in part (b) has almost exactly the shape of the excited-state wave function ψ_2 .

*As we will see in the next example, there is also a small probability of finding the electron in certain higher levels, $m > 2$, but these probabilities are so much smaller than that for $m = 2$ that the form (11.29) is an excellent approximation.



346 Chapter 11 • Atomic Transitions and Radiation

In general, the effect of switching on a perturbation W is to add to an initial wave function ψ_n the extra term of (11.27) proportional to $W\psi_n$, which may contain small components of several different wave functions ψ_m . For each $m \neq n$, there is therefore a definite probability $P(n \rightarrow m)$ that a system that started out in the state ψ_n will be found, after the perturbation has acted, in the state ψ_m . To evaluate this probability, we can expand the extra term as in (11.20)

$$-\frac{i \Delta t}{\hbar} W\psi_n = \sum_m A_m \psi_m \quad (11.30)$$

and according to (11.24), the coefficient A_m is given by the integral

$$A_m = -\frac{i \Delta t}{\hbar} \int_{-\infty}^{\infty} \psi_m^* W \psi_n dx. \quad (11.31)$$

The **transition probability** $P(n \rightarrow m)$ is then just $|A_m|^2$, so

$$P(n \rightarrow m) = \left(\frac{\Delta t}{\hbar} \right)^2 \left| \int_{-\infty}^{\infty} \psi_m^* W \psi_n dx \right|^2 \quad (11.32)$$

The integral inside the absolute value signs is called the **matrix element** for the transition $(n \rightarrow m)$. By evaluating this integral, we can find the probability for any transition in which we may be interested.

Example 11.4

Consider again the electron of Example 11.3, which is initially in the ground state of the infinite square well and is exposed to an electric field \mathcal{E} for a short time Δt . Find the probabilities $P(1 \rightarrow m)$ that it will subsequently be found in the level m for $m = 2, 3$, and 4 .

The required probabilities are given by (11.32), with $W = e\mathcal{E}x$. Thus,

$$P(n \rightarrow m) = \left(\frac{e\mathcal{E} \Delta t}{\hbar} \right)^2 \left| \int_{-\infty}^{\infty} \psi_m^* x \psi_n dx \right|^2 \quad (11.33)$$

The evaluation of these integrals is a straightforward (though surprisingly tedious) exercise. The wave functions ψ_n are given in (11.21), although we have to rewrite them since we have moved our origin to the center of the well. You should have no difficulty in showing that, with the well centered on the origin, the wave functions with n odd become

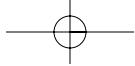
$$\psi_n = \sqrt{\frac{2}{a}} \cos \frac{n\pi x}{a} \quad (n \text{ odd}) \quad (11.34)$$

whereas, for n even, they become

$$\psi_n = \sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \quad (n \text{ even}) \quad (11.35)$$

Thus, the integral of (11.33) for the case that $n = 1$ and $m = 2$, is

$$\int_{-\infty}^{\infty} \psi_2^* x \psi_1 dx = \frac{2}{a} \int_{-a/2}^{a/2} \sin \frac{2\pi x}{a} x \cos \frac{\pi x}{a} dx \quad (11.36)$$

Section 11.5 • Transitions; Time-Dependent Perturbation Theory **347**

which can be evaluated (Problem 11.21) to give $8a/9\pi^2$. Thus,

$$P(1 \rightarrow 2) = \left(\frac{e\mathcal{E}\Delta t}{\hbar} \frac{8a}{9\pi^2} \right)^2 \quad (11.37)$$

In exactly the same way, we can evaluate

$$P(1 \rightarrow 4) = \left(\frac{e\mathcal{E}\Delta t}{\hbar} \frac{16a}{225\pi^2} \right)^2,$$

some 150 times smaller than $P(1 \rightarrow 2)$. The transition $1 \rightarrow 3$ is more interesting: The integral of (11.33) is

$$\int_{-\infty}^{\infty} \psi_3^* x \psi_1 dx = \frac{2}{a} \int_{-a/2}^{a/2} \cos \frac{3\pi x}{a} x \cos \frac{\pi x}{a} dx \quad (11.38)$$

There is no need to evaluate this integral: It is easy to see that the integrand is an odd function of x ; that is, its value at $-x$ is the negative of its value at x . Since the range of integration is the symmetric interval $-a/2$ to $a/2$, the contribution from any point x is cancelled by that from $-x$, and the integral is exactly zero. Thus, at least within our approximations, the probability for the transition $1 \rightarrow 3$ is zero, $P(1 \rightarrow 3) = 0$.

It is easy to see that this last result applies to any transition $n \rightarrow m$ with both n and m odd, and likewise with both n and m even. Because of the symmetries of the sine and cosine functions, the integrals concerned are zero, and

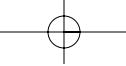
$$P(\text{odd} \rightarrow \text{odd}) = P(\text{even} \rightarrow \text{even}) = 0 \quad (11.39)$$

Selection Rules

The final result of the last example is an instance of an important phenomenon. It often happens that because of the symmetries of the wave functions, the integral involved in the transition probability $P(n \rightarrow m)$ is exactly zero. When this happens, we say that the transition concerned is a **forbidden transition**, and within the approximations used the probability $P(n \rightarrow m)$ is zero. Strictly speaking, this does not mean that the transition is impossible, since the expression (11.32) is only an approximation. Nevertheless, it is usually the case that the probability of forbidden transitions is small compared to other “non-forbidden” transitions. With this terminology, we can say that in the last example, transitions between any two odd levels (and likewise between any two even levels) are forbidden. A rule of this kind, specifying a whole family of forbidden transitions, is called a **selection rule**. Selection rules are very useful in deciding which transitions of a quantum system should be observable with appreciable probability.

Upward and Downward Transitions

In our example of the electron in the infinite square well, we assumed that the system started out in its ground state, so that all other states had higher energy, and the only possible transitions were upward transitions to a higher level. However, the result (11.32) applies equally whether the level n is lower than the level m or vice versa. Thus, if the electron were initially in the level $n = 2$,



348 Chapter 11 • Atomic Transitions and Radiation

switching on an electric field could cause it to drop downward to the ground state. Transitions from a higher to a lower level usually result in the emission of a photon. Thus, a downward transition caused by an externally applied field is called stimulated emission — a process that plays a key role in the operation of lasers, as we discuss shortly.

Perturbations that Last a Long Time

So far we have considered only a perturbation that is switched on for a short time Δt . It often happens that the external influence that causes atomic transitions lasts for a long time. For example, we could place the atom in a steady beam of light, in which case the atom would feel oscillating electric and magnetic fields that last for hours or even days. To handle this, we need to extend our time-dependent perturbation theory to cover longer time intervals. This extension is unfortunately beyond our scope here, but the main results can be easily described. It turns out that the transition probability is given by a formula very similar to (11.32) except that (as you might guess) it involves an integral over time. In particular, it remains true as in (11.32) that the transition probability is proportional to the absolute value squared of a “matrix element”

$$P(n \rightarrow m) \propto \left| \int_{-\infty}^{\infty} \psi_m^* W \psi_n dx \right|^2 \quad (11.40)$$

Thus, many of the ideas discussed above still apply. For example, when symmetries of the perturbation W and the wave functions imply that the integral in (11.40) is zero, we say that we have a selection rule, and the transition concerned is said to be forbidden.

An important example of a perturbation that lasts a long time is the oscillating electric field associated with electromagnetic radiation that we could direct at an atom. In this case the perturbation would have the form $e\epsilon_0 x \cos \omega t$, or, a little more generally,

$$W(\mathbf{r}) \cos \omega t \quad (11.41)$$

where ω is the angular frequency of the applied radiation. In this case the expression inside the absolute values signs of (11.40) must include a time integral as well. This integral over time turns out to be negligibly small unless the frequency ω of the applied radiation is close to the value

$$\omega = |E_m - E_n|/\hbar \quad (11.42)$$

That is, transitions occur appreciably only if $\hbar\omega$ is close to the energy difference between the two levels concerned. This result agrees with our knowledge that a photon carrying energy $\hbar\omega$ can cause an upward transition ($n \rightarrow m$) with $E_m > E_n$ only if $\hbar\omega$ is equal to the difference $E_m - E_n$, in which case the photon is absorbed. However, there is a second important possibility: If $E_m < E_n$, the *downward* transition $n \rightarrow m$ can occur, provided $\hbar\omega$ is close to $E_n - E_m$. In this case, energy conservation requires that the atom release energy, which it does by emitting a photon of energy $\hbar\omega = E_n - E_m$, so that the incident radiation gains one photon. Since the emission of this photon is caused, or *stimulated*, by the applied radiation, it is called **stimulated emission of radiation**. Both of these possibilities are illustrated schematically in Fig. 11.3. The possibility of stimulated emission, depicted in part (b), is the basis of the laser, as we discuss in Sections 11.9 and 11.10.

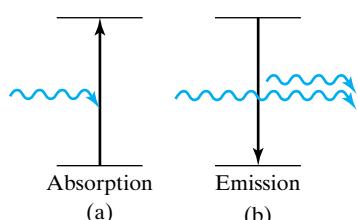


FIGURE 11.3

Schematic diagram of absorption and stimulated emission. (a) An incident photon is absorbed by the atom, which makes an upward transition. (b) In stimulated emission one photon striking the excited atom stimulates the emission of a second photon, causing the atom to make a downward transition.

11.6 A Brief Review

Let us review briefly the results of the previous section. If you read that section, this review may help to reinforce its ideas, and if you didn't, this will establish the ideas you need to take on faith. The central result is this: Suppose that an isolated quantum system is initially in one of its stationary states ψ_n . Suppose next that the system is subjected to some outside perturbation giving it an additional potential energy, which could be a static $W(\mathbf{r})$ or an oscillatory $W(\mathbf{r}) \cos \omega t$. Then, at any later time, there is a definite probability $P(n \rightarrow m)$ that the system, which was in the state ψ_n , will be found in a different state ψ_m . This transition probability can be calculated by the method of time-dependent perturbation theory and is found to have the form (11.40)

$$P(n \rightarrow m) \propto \left| \int \psi_m^* W \psi_n dV \right|^2 \quad (11.43)$$

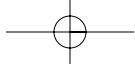
written here in the form appropriate for an electron in three dimensions, with $dV = dx dy dz$. If the perturbation is the oscillating electric field of electromagnetic radiation, the transition $(n \rightarrow m)$ can be an upward transition, resulting in the absorption of one of the incident photons, or a downward transition, resulting in the **stimulated emission** of another photon of the same frequency. In either case transitions associated with electromagnetic radiation and involving the emission or absorption of photons are called **radiative transitions**.

It often happens that the integral (or “matrix element”) in (11.43) can be shown to be zero on the grounds of symmetries of the perturbation W and the two wave functions. In this case we say that the transition $(n \rightarrow m)$ is **forbidden**, although this usually means only that the transition is much less probable than other “non-forbidden” transitions.

11.7 Spontaneous Emission

We have seen that radiation directed at an excited atom can cause it to make a downward transition, emitting a photon as it does so. However, we also know that an excited atom can make a downward transition even when it is isolated from any applied radiation. This process is called **spontaneous emission**. As we discussed in Section 11.3, an isolated atom in a truly stationary state, excited or not, would remain forever in that state. How, then, do excited atoms give off light without the aid of stimulating radiation? The answer is that the radiation required to stimulate downward transitions is *always* present, even when an atom is perfectly isolated from all apparent sources. The origin of this ubiquitous radiation has to do with the quantization of the electromagnetic field.

We know that the classical description of the electromagnetic field fails to explain the quantization of radiation. The complete quantum treatment of electromagnetic radiation is called **quantum electrodynamics** (or **QED**) and is beyond the scope of this book. However, we can easily describe its predictions relevant to our present discussion: At any given frequency, ω , the energy of radiation is quantized and can change only by integer multiples of $\hbar\omega$, corresponding to the now-familiar concept of photons. Furthermore, QED predicts that the lowest possible energy of the radiation field is *not* zero. That is, when one treats the electromagnetic field as a quantum system, its lowest energy level is found to be greater than zero.



350 Chapter 11 • Atomic Transitions and Radiation

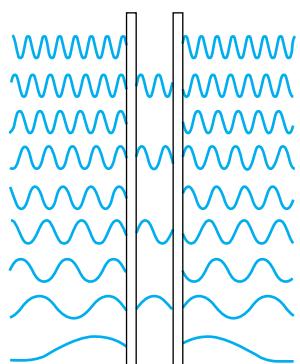


FIGURE 11.4

In the Casimir effect, zero-point radiation exerts forces on two uncharged parallel metal plates. Outside the plates, all wavelengths of radiation are allowed; between the plates, only certain discrete wavelengths are allowed (just as only certain discrete wavelengths are allowed on a stretched string). If the plates are close enough together, this difference produces a measurable inward force, due to the unequal radiation on opposite sides of either plate.

We have seen similar behavior in other quantum systems. For instance, the minimum energy of a particle in a box is nonzero. As we saw in Section 6.8, the reason for this is the Heisenberg uncertainty relation between position and momentum. A similar uncertainty relation applies to the electric and magnetic fields \mathcal{E} and B and implies that they cannot both be exactly zero, even in empty space. As a result, the electromagnetic field has a nonzero amplitude at every frequency. This minimum possible field is called the **zero point**, or **vacuum fluctuation, field**.

In many situations the effects of the zero-point fields are negligible. Nevertheless, the presence of these fields can be demonstrated in an experiment first proposed by the Dutch physicist Hendrik Casimir. As illustrated in Fig. 11.4, this experiment measures a tiny force pushing two uncharged parallel metal plates together. This force would not exist if space were not filled by the zero-point fields.

For our present purposes, a far more important consequence of the zero-point fields is their effect on an isolated atom. If the isolated atom is in an excited state, the zero-point fields can stimulate it to emit a photon and drop to a lower state. Since this process occurs without externally applied radiation, it is called spontaneous emission. However, you should recognize that stimulated and spontaneous emission are essentially the same phenomenon. Stimulated emission is caused by an externally applied field. Spontaneous emission is just stimulated emission at its lowest possible rate, which occurs when the electromagnetic fields are at their minimum possible (zero-point) level.

Although the zero-point fields cause spontaneous downward transitions, conservation of energy does not allow them to cause upward transitions. For an upward transition to occur, the energy gained by the atom must be balanced by a loss of energy from the radiation field. This is clearly impossible when the energy of the radiation field is already at its lowest possible value. In the special case of an atom in its ground state, this means that there can be no spontaneous transitions at all. Upward transitions cannot occur because energy could not be conserved, and downward transitions do not occur because there are no lower states. Thus the ground state of an isolated atom is truly stable.

11.8 Atomic Selection Rules

Let us summarize our findings as they apply to an electron in an atom exposed to radiation (either radiation that is applied externally or just that of the zero-point fields). The most important effect of the radiation comes from the oscillatory electric field of the form $\mathcal{E} = \mathcal{E}_0 \cos \omega t$. (Of course, there is also an oscillatory magnetic field, but its effect is usually much less important.) Thus, the dominant perturbation has the form

$$e\mathcal{E}x = e\mathcal{E}_0x \cos \omega t \quad (11.44)$$

(This is if the electric field points in the x direction. In three dimensions, it can, of course point in any direction.) If the electron is initially in a state ψ_n , then the probability $P(n \rightarrow m)$ that at a later time it will be found in a state ψ_m is given by (11.43) as

$$P(n \rightarrow m) \propto \mathcal{E}_0^2 \left| \int \psi_m^* x \psi_n dV \right|^2 \quad (11.45)$$

This gives the probability of the transition ($n \rightarrow m$) if the perturbing electric field is in the x direction. If the field is in the y direction, we must simply replace x by y , and so on. And if the perturbing radiation is unpolarized (as is certainly the case for the zero-point radiation on an otherwise isolated atom), we must average over all three directions.

As we mentioned earlier, if the matrix element in (11.45) is zero, we say that the transition ($n \rightarrow m$) is forbidden, meaning that at least within the approximations that led to (11.45) the probability $P(n \rightarrow m)$ is zero. It is sometimes surprisingly easy to show that this is the case. For example, suppose that the initial state ψ_n is an s state (with $l = 0$) and suppose that we are interested in the possibility of a transition to another s state ψ_m . In this case both of the wave functions in (11.45) are spherically symmetric. Therefore, the contribution to the integral from any point (x, y, z) is exactly cancelled by the contribution from the point $(-x, y, z)$, and the whole integral is automatically zero. That is, all transitions from one s state to another are forbidden (meaning in practice that such transitions are anyway extremely improbable). This is a particular case of a more general **selection rule** that any transition ($n \rightarrow m$) is forbidden unless $|l_m - l_n| = \pm 1$, a rule that we usually state thus:

$$\text{any transition } (n \rightarrow m) \text{ is forbidden unless } \Delta l = \pm 1 \quad (11.46)$$

The proof of this selection rule is beyond our scope here, but is illustrated by some simple cases in Problems 11.26, 11.30, and 11.31. The rule is easy to understand if we recall that the photon has spin 1. (That is, it has an intrinsic angular momentum of magnitude $\sqrt{s(s+1)}\hbar$ with $s = 1$.) This means that when a photon is emitted by or absorbed into an atom, it must change the atom's angular momentum by 1 unit, as in (11.46). (This is actually a quite subtle argument — see Problem 11.31.)

The selection rule (11.46) is illustrated in Fig. 11.5, which shows some of the allowed (that is, non-forbidden) transitions in the hydrogen atom. While there are many such allowed transitions, there are even more that are forbidden. For example, all transitions of the form $s \rightarrow s$ are forbidden (as we proved above) and so are any of the form $s \leftrightarrow d$ or $s \leftrightarrow f$. Similarly, both $p \rightarrow p$ and $p \leftrightarrow f$ are forbidden. We can express all this compactly by saying that a state in one column of the energy-level diagram can make transitions only to states in one of the adjacent columns. In particular, s states can make transitions only to p states. Other examples of selection rules appear in Problems 11.25, 11.27,

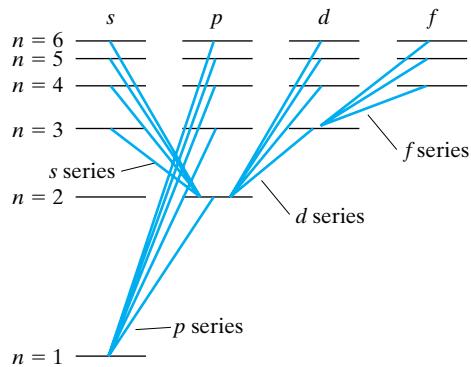
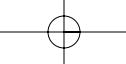


FIGURE 11.5

Some of the allowed transitions observed in the hydrogen atom. Note that each involves a change of l by one unit, as is found to be the case for all allowed transitions. Note also that the traditional labels s (sharp), p (principal), d (diffuse), and f (fundamental) were originally applied to transitions, not levels.



352 Chapter 11 • Atomic Transitions and Radiation

and 11.30. For reference, we have collected together some of the more important selection rules in Table 11.1.

TABLE 11.1

Some of the selection rules that apply to transitions of electrons in an atom. Each rule is stated in the form of a condition that must be met if a transition is to be allowed (that is, occur with significant probability). For example, the first rule, $\Delta l = \pm 1$, means that only transitions for which $\Delta l = l_f - l_i = \pm 1$ are allowed. The quantum number s_{tot} identifies the magnitude of the total spin of all the electrons; similarly j_{tot} gives the magnitude of the total angular momentum $\sum(\mathbf{L} + \mathbf{S})$.

Quantum Number	Selection Rule	Reference
l (magnitude of \mathbf{L})	$\Delta l = \pm 1$	Eq. (11.46)
m (z component of \mathbf{L})	$\Delta m = 0$ or ± 1	Problem 11.30
s_{tot} (total spin $\sum \mathbf{S}$)	$\Delta s_{\text{tot}} = 0$	Problem 11.25
j_{tot} [total spin + orbital $\sum(\mathbf{L} + \mathbf{S})$]	$\Delta j_{\text{tot}} = 0$ or ± 1	Problem 11.27

Metastable States

Looking back at Fig. 11.5, you can see that the $2s$ level of hydrogen has no allowed downward transitions because there is no $l = 1$ level below the $2s$ level. This would seem to imply that the $2s$ state is perfectly stable. In fact, there exist other processes that de-excite the $2s$ state, such as a collision with other atoms.* However, these processes all occur very slowly, and the lifetime of the $2s$ state is exceptionally long. There are many other atoms that have similar excited states with no allowed downward transitions. When this occurs the long-lived excited state is called a **metastable state**. These states are important in the operation of many lasers, as we see in the next section.

11.9 Lasers

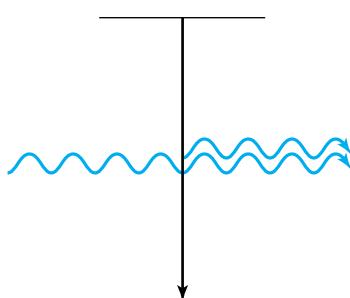


FIGURE 11.6

Incident radiation stimulates a transition from an excited state to a lower state. The emitted radiation is exactly in phase with the incident radiation.

Lasers are made possible by the abstruse-seeming phenomena of stimulated emission and metastable states. Today lasers play such a large role in our lives it is hard to imagine life without them. In the last two sections of this chapter we will describe how several kinds of laser work and discuss briefly some of their uses.

The process of stimulated emission is illustrated in Fig. 11.6. A single photon of the correct frequency can stimulate an excited atom to drop to a lower level and emit another photon of the same frequency. If there are many excited atoms present, each of these two photons can go on to stimulate two more emissions, producing four photons in all. As long as the majority of atoms are still in the excited state, this process can continue in a cascade, giving

$$1 \text{ photon} \rightarrow 2 \text{ photons} \rightarrow 4 \text{ photons} \rightarrow 8 \text{ photons} \rightarrow \dots$$

and so on. In other words, the process of stimulated emission can produce a dramatic *amplification* of a beam of photons. The **laser** is a device that exploits

*Another such process, which occurs even in isolated atoms, is the simultaneous emission of two or more photons, but this occurs extremely rarely.

this possibility to amplify light of a definite frequency. The name laser is an acronym for "Light Amplification by the Stimulated Emission of Radiation."

In stimulated emission the time dependence of the oscillation induced in each atom is in lockstep with the stimulating radiation. This means that the emitted radiation is exactly in phase with the radiation causing the transition, as suggested in Fig. 11.6. This **coherence** of the emitted photons means that the resulting light wave is an almost perfect sinusoidal wave. This contrasts with the **incoherent** light from ordinary sources where the phase of each atom's radiation is random with respect to all the others. It can also be shown that this coherence means that the laser beam can have a very well defined direction in space.

We have described the amplification of light in a laser as caused by stimulated emission from excited atoms. Normally, however, most atoms are in their ground state, and radiation that strikes atoms in their ground state will be absorbed, not amplified. To achieve amplification, we must arrange that a majority of the atoms are in an excited state. This reversal of the normal population of the levels is called a **population inversion**. Different lasers use different means to achieve population inversion, as we describe in the following accounts of some important types of lasers. Some lasers produce short pulses of light, while others produce a continuous beam. Since the details of operation of these two types are somewhat different, we describe them in turn, starting with the pulsed laser.

Pulsed Lasers

The first successful laser for visible light* was a pulsed laser, developed by the American physicist Theodore Maiman in 1960. The essential element of this laser was a ruby rod containing chromium ions, which have a metastable state 1.79 eV above the ground state. Transitions between these two states produce photons with a wavelength $\lambda = 694 \text{ nm}$, in the deep red portion of the visible spectrum.

The general design of the ruby laser is shown in Fig. 11.7. In the first step of operation, a brilliant burst of light from a flash lamp causes most of the Cr

Charles Townes (born 1915, American)



The principle of the laser was first developed for microwaves, the corresponding device being the maser (an acronym for Microwave Amplification by Stimulated Emission of Radiation). Townes had a large part — both theoretical and practical — in development of the maser. He also contributed substantially to the laser (which amplifies light). Both masers and lasers have revolutionized almost every branch of science; for example, the maser is the basis of the atomic clock. For his role in all this, Townes shared the 1964 Nobel Prize in physics. In the photo, Townes is standing next to a maser amplifier for radio Astronomy.

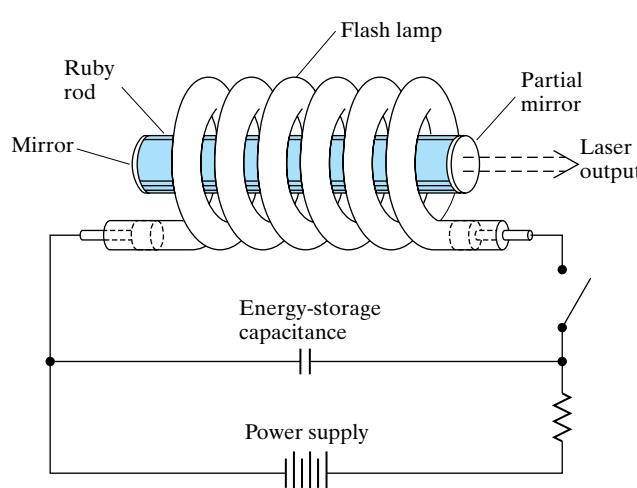
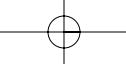


FIGURE 11.7

Schematic drawing of a pulsed ruby laser. The flash lamp excites Cr ions in the ruby rod. The mirrors cause light to reflect back and forth in the rod to increase the probability of stimulating further emission. One mirror is not 100% reflective, so some light escapes to form the external beam.

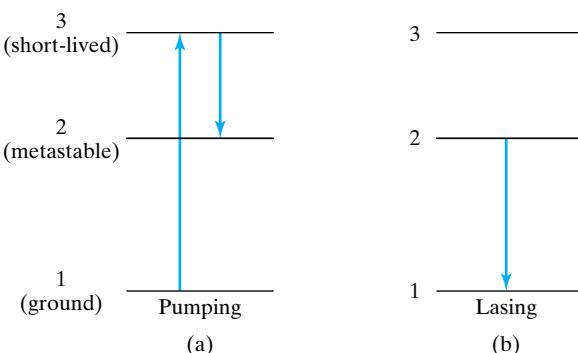
*We should mention that the first use of stimulated emission to amplify any kind of radiation was with microwaves. Devices that amplify microwaves in this way are called masers.



354 Chapter 11 • Atomic Transitions and Radiation

FIGURE 11.8

The relevant levels of the chromium ions in a ruby laser are: (1) the ground state, (2) the metastable state at 1.79 eV, and (3) a short-lived state. (a) Absorption of flash-lamp light elevates many Cr ions to the short-lived state, which promptly decays to the metastable state, where a large population builds up. (b) The process of stimulated emission then causes the laser transition to occur in many Cr ions.



Theodore Maiman
(born 1927, American)



After getting his BA from the University of Colorado and his PhD from Stanford, Maiman joined the Hughes Research Laboratory in Miami, where he developed the first working laser in 1960. He subsequently founded several independent companies, one of which, Korad Corporation, became the leading developer and manufacturer of high-power lasers.

ions in the ruby rod to be excited out of their ground state. The majority of these excitations populate a short-lived state that makes a prompt transition to the metastable state at 1.79 eV, as shown in Fig. 11.8. (The process of moving the ions from their ground state to the metastable level is called **pumping**.) Because the metastable level is relatively long-lived (about 4 ms) a majority of the Cr ions are caught, briefly, in this level. Once the number of ions in the metastable state is greater than the number in the ground state ($N_2 > N_1$), any light produced by a spontaneous $2 \rightarrow 1$ transition is amplified by stimulated emission from the excited ions it encounters. As this amplified light sweeps through the rod, it rapidly de-excites many excited ions and forms an intense, short pulse of light. This cascade can occur only if there is a population inversion ($N_2 > N_1$), for if $N_2 \leq N_1$, then absorption exceeds or equals emission, killing the amplification process. When the majority of the ions have returned to their ground state, the entire process can be repeated.

To give the light ample opportunity to stimulate emission, it is reflected back and forth along the rod, whose ends are polished and silvered to form two mirrors.* One of these mirrors is made partially transparent to allow light to escape and form the external beam of laser light. In the original laser the xenon flash lamp was powered by a capacitor bank of several hundred microfarads charged to several kilovolts. The principal parts of this laser can be seen in the photograph in Fig. 11.9.

There are three major differences between the pumping light from the flash lamp and the light produced by the laser action itself. First, the flash lamp emits a broad range of wavelengths (white light), while the laser light is concentrated in a narrow spectral line at $\lambda = 694$ nm. Second, the flash lamp light is incoherent, while the laser light is coherent. Third, the flash lamp light radiates in all directions, whereas the laser beam can be designed to have a narrowly defined direction in space — an important property in many applications.

Since the development of the ruby laser, a variety of other materials have been found that can be pumped to a metastable level and used to make pulsed lasers. These lasers vary widely as to total energy and pulse length. For the ruby laser, a representative value for the energy of a single pulse is on the order of 10 J. Since each pulse has a typical duration of order[†] $100 \mu\text{s}$, this gives an instantaneous power of order 100 kW. This high instantaneous power is the basis of many laser applications: Small pulsed lasers are used in medicine

*The distance L between the mirrors must satisfy the condition $L = n\lambda/2$, to ensure constructive interference of the multiply reflected waves.

[†] As we discuss in Section 11.10 this pulse is really a cluster of much shorter pulses.

**FIGURE 11.9**

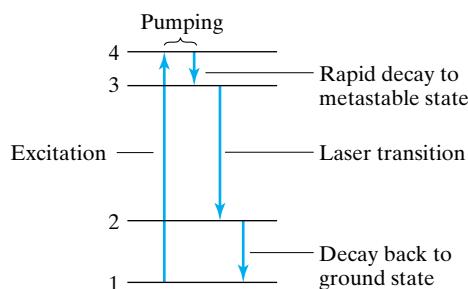
The original laser, built by Maiman at the Hughes Research Lab. The ruby rod, about 1 cm. in diameter, can be seen inside the coiled flash lamp.

to cause coagulation and to suture tissues by forming tiny scars. Larger lasers are widely used in industry for welding, perforating, and machining. Yet larger versions are being investigated for military use.

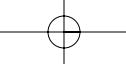
Continuous Lasers

For many purposes, it is convenient to have a continuous wave of laser light. Most **continuous-wave** (or **CW**) lasers use *four* levels, instead of the three levels used in the ruby laser. This avoids an important problem with any three-level laser: In a three-level laser, more than half the atoms must be excited out of their ground state to achieve the necessary population inversion, with $N_2 > N_1$. Since nearly all atoms are normally in the ground state, a great deal of energy must be supplied by the pumping flash to accomplish this. A **four-level laser** greatly reduces this requirement by making the lower level of the laser transition an excited state which is normally empty. As shown in Fig. 11.10, the laser transition from level 3 to level 2 avoids the heavily populated ground state so that the condition of population inversion ($N_3 > N_2$ in this case) is more easily achieved. Because of their lower power requirements, such lasers can operate continuously.

A popular type of continuous four-level laser is the helium–neon laser. This uses a mixture of helium and neon gases and produces red light with $\lambda = 633$ nm. The uppermost level of this laser is in the helium atoms, and the other three are in the neon, as shown in Fig. 11.11. The He atoms are pumped by high-speed electrons rather than by a flash lamp. This is achieved by a

**FIGURE 11.10**

In a four-level laser the laser transition is between levels 3 and 2. Since few atoms are in level 2, the condition $N_3 > N_2$ is readily achieved.



356 Chapter 11 • Atomic Transitions and Radiation

FIGURE 11.11

The level initially pumped in the He–Ne laser is in the He atoms. Collisions transfer this energy to a level in the Ne atoms, which then produce stimulated emission, terminating in a nearly empty excited state.

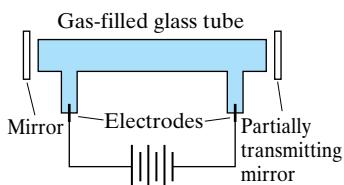
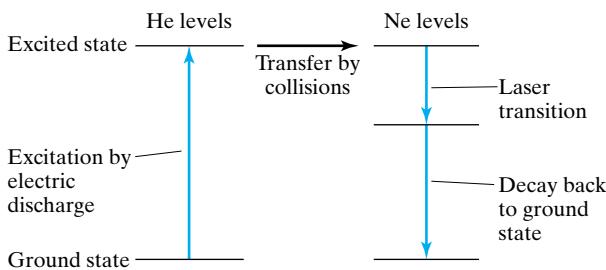


FIGURE 11.12

Key components of an He–Ne laser are the gas-filled glass tube, mirrors, and power supply.

steady electric discharge (like that in a neon sign) within the He–Ne mixture. The energetic electrons in the discharge strongly excite a metastable level in He whose energy is very close to one of the excited levels of Ne. In collisions between the excited He atoms and unexcited Ne atoms, there is a high probability for transfer of the excitation energy, with the He atom dropping to its ground state while the Ne atom is excited. (The upper level in Ne is not so readily excited directly by fast electrons, hence the need for He and the transfer process.) The laser transition occurs in the Ne, producing the characteristic red light at 633 nm.

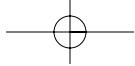
Not many atoms are required to be in the uppermost Ne level because the level below it rapidly empties by a fast transition to the ground state. The condition $N_3 > N_2$ is therefore achieved with only modest power requirements for the electric discharge.

Figure 11.12 is a schematic of an He–Ne laser. The gas mixture fills a glass tube fitted with electrodes to produce a continuous electric discharge. As in the ruby laser, mirrors reflect light back and forth to increase the opportunity for stimulated emission. The external beam passes out through one of the mirrors, which is partially transmitting. Typical small models produce a light beam with a power of 10^{-3} W and consume a few watts of electric power.

Applications

The advent of the laser has totally transformed research in many branches of science. The sharply defined frequency of laser light has made possible measurements of quantized energy levels in atoms, molecules, and solids, with unprecedented accuracy. Because laser light is so intense, measurements can be made with very small samples — transitions in *single* atoms have even been observed. Using lasers with pulses that last just a few femtoseconds ($1 \text{ fs} = 10^{-15} \text{ s}$), chemists can follow the detailed evolution of chemical reactions. The momentum of the photons in a laser beam can be used to cool a gas to temperatures of order 10^{-6} kelvin (Chapter 13), but also to confine the plasma in a fusion device at higher than 10^6 kelvin (Chapter 17).

Lasers also allow extremely accurate measurements of distances. For example, in the lunar ranging experiment, the distance from the earth to the moon is measured by timing the round trip of a laser pulse that is fired at the moon and reflected back by a mirror placed there by Apollo astronauts; in this way changes in the earth–moon distance of a few centimeters can be detected. Similarly, geophysicists can monitor the extremely slow motions of the tectonic plates that comprise the earth’s crust. Laser interferometers (described in Section 11.10) allow length measurements with accuracies of order one hundredth of the wavelength of light (that is, accuracies of a few nanometers). With the help of lasers, the speed of light can be measured so accurately that it has now become a defined constant, in terms of which the



SI meter is specified. And lasers play an essential role in the atomic clocks on which our timekeeping system (an essential part of the global positioning system) depends.

We have already mentioned some commercial applications of lasers, and it is easy to list more — in communications using optical fibers, in sound recording using compact discs, in computing, and so on. In medicine, lasers are used for cancer therapy and for welding tissue (in the human retina for example). To conclude, we mention just one more application, the laser gyroscope, which has replaced the conventional gyroscope in many commercial aircraft and is described in Fig. 11.13.



11.10 Further Properties of Lasers*

*This section can be omitted without significant loss of continuity.

Angular Divergence

The directional properties of laser beams depend on the paths followed by light within the laser cavity (the space between the two mirrors). Figure 11.14(a) shows light reverberating back and forth in a straight line between the mirrors as its amplitude is increased by stimulated emission from excited atoms. A laser operating in this manner is called a single-mode laser. If the laser medium were perfectly homogeneous and the mirrors perfectly flat, the external beam from single-mode operation would be nondiverging if light were not a wave. However, because of the wave nature of light, diffraction occurs and causes the light to diverge slightly. The minimum angle of divergence, $\delta\theta$ (defined as the angle from center to first minimum), depends on the diameter d of the beam and is given by the diffraction relation*

$$\delta\theta \approx \frac{\lambda}{d} \quad (11.47)$$

where λ is the wavelength of light (and $\delta\theta$ is measured in radians). With typical laser dimensions, the actual angle of divergence $\delta\theta$ is of order 10^{-4} rad, which is extremely small compared with the spreading from most other light sources. Nevertheless, over large distances, even laser beams suffer appreciable spreading, which reduces their intensity, as we see in the following example.

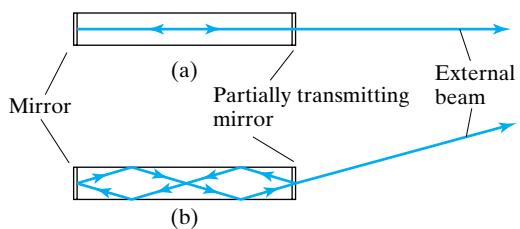


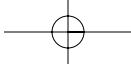
FIGURE 11.14

(a) A laser operating on only one mode, in which light reflects back and forth along a straight path between the mirrors. (b) A more complicated mode, which produces light traveling at an (exaggerated) angle.

*For a circular aperture, this formula is generally quoted as $\delta\theta = 1.22 \lambda/d$. For our purposes of estimation, we have dropped the factor of 1.22.

FIGURE 11.13

A laser gyroscope, consisting of three perpendicular, ring-shaped lasers, all bored in a single 5-inch cube of glass. Each ring contains two laser beams, rotating in opposite directions. The interference between the two beams is extremely sensitive to rotations of the ring, and with three such rings, the device can be used like an ordinary gyroscope to keep track of an aircraft's orientation.



358 Chapter 11 • Atomic Transitions and Radiation

Example 11.5

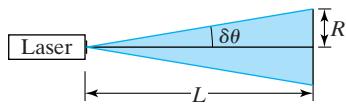


FIGURE 11.15

The radius of a laser beam at a distance L is approximately $R \approx L \delta\theta$. (Because R is much larger than the original beam size, we can neglect the latter.)

A single-mode He-Ne laser ($\lambda = 633$ nm) with an initial beam diameter of 3 mm has an angular divergence limited only by diffraction. Find the approximate diameter of the beam at a distance of 300 m. By what factor is the intensity reduced at this distance?

According to (11.47), $\delta\theta \approx \lambda/d$; therefore, as shown in Fig. 11.15, the radius R of the beam at a distance L is

$$R \approx L \delta\theta \approx \frac{L\lambda}{d}$$

With $L = 300$ m, this gives

$$R = \frac{(300 \text{ m}) \times (6.3 \times 10^{-7} \text{ m})}{3 \times 10^{-3} \text{ m}} \approx 60 \text{ mm}$$

Therefore, the beam diameter is about 120 mm.

Since intensity is power per unit area, it is inversely proportional to the cross-sectional area of the beam. Thus the ratio of final to initial intensities is

$$\frac{I_f}{I_i} = \frac{\pi d_i^2 / 4}{\pi d_f^2 / 4} = \left(\frac{d_i}{d_f} \right)^2 \approx \frac{1}{1600}$$

This loss of intensity is sufficiently large that it must obviously be taken into account when planning to use lasers over large distances.

More complex modes, like that in Fig. 11.14(b), can also occur. These modes produce nonaxial light that increases the angular divergence of the laser beam; typically, $\delta\theta \approx 10^{-3}$ rad. A small aperture can be inserted in the laser cavity to limit laser action to the axial mode and hence reduce the angular divergence, but this usually results in a loss of total power.

Time Dependence of Pulsed Laser Beams

In a pulsed laser the flash lamp that initiates the pulse usually produces its white light over a period of several hundred microseconds, as shown in Fig. 11.16(a). The output of the laser generally consists of a series of spikes, as in Fig. 11.16(b). The occurrence of these brief spikes is easily understood. Initially, in the first 100 μs or so, $N_2 < N_1$, so laser action cannot occur. Only spontaneous emission, which is incoherent, occurs during this time. When N_2

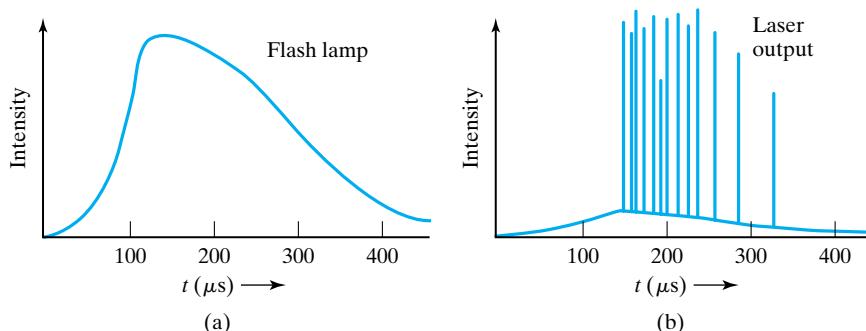


FIGURE 11.16

(a) The flash-lamp intensity in a pulsed laser typically lasts several hundred μs . (b) The output of the laser contains a weak background of incoherent spontaneous emission and a series of intense spikes of coherent light.

finally exceeds N_1 , laser amplification begins and the light level within the cavity builds rapidly. In fact, a “runaway” occurs because the rate of stimulated emission grows as the amplitude of the radiation increases. Very quickly, then, N_2 decreases until $N_2 < N_1$, and laser amplification ceases, typically in a fraction of a microsecond. However, the pumping radiation from the flash lamp is still present and again causes N_2 to increase until $N_2 > N_1$, producing another spike. This process continues until the level of pumping radiation falls below that which can make N_2 greater than N_1 .

Q-Switching

The size of the pulses produced by a ruby laser is limited because laser amplification begins as soon as $N_2 > N_1$ and ends quickly when N_2 drops back to $N_2 \approx N_1$. Larger pulses could be produced if N_2 could be greatly increased before laser amplification begins. This can be accomplished by preventing amplification temporarily until N_2 is large and then permitting laser action to occur. This method of producing single, very intense pulses is called **Q-switching**. The name originates in the term **quality factor**, generally abbreviated as Q . This parameter is a measure of the time light can reverberate within the laser cavity before it dies away.* If Q is large, the light loses energy (by scattering, for example) slowly and hence is easily able to stimulate laser amplification. If Q is small, the light loses energy quickly, and if Q is sufficiently small, laser action does not occur.

In the lasers described so far, the cavity is designed with high Q by using low-loss mirrors and a clear laser medium so that laser amplification is easily achieved. This causes laser action to begin very soon after $N_2 > N_1$ and limits the size of the pulse. However, if the Q of the cavity is reduced to a low level, laser action cannot occur even with all the atoms in the excited state. A very large pulse can then be produced by abruptly increasing Q (hence the name “ Q -switching”).

A variety of schemes are used for the temporary reduction of Q . The simplest method to understand uses a rotating mirror for one of the cavity mirrors as sketched in Fig. 11.17. The laser cannot operate until the mirrors are parallel. The flash lamp is fired a few hundred microseconds before the mirrors become parallel, so that the metastable level becomes highly populated. The pulse that occurs when the mirrors become parallel rapidly drains the population of level 2, producing a single extremely intense pulse. Figure 11.17(b) shows the pulse produced by a laser of this type with a peak power of 100 MW. (The peak power of such lasers can be as large as thousands of megawatts.)

Another Q -switching technique uses a dye contained in a transparent cell that is placed between the ruby rod and one of the mirrors. The color of a dye is due to its absorption of light at wavelengths corresponding to the energy difference between its ground state and a group, or *band*, of excited states. By use of a dye that absorbs strongly at the laser wavelength, Q is reduced so much that laser action cannot occur. However, after the flash lamp has operated for some time, virtually all of the dye molecules have been excited out of their ground state, so that absorption ceases.[†] At this point the dye becomes transparent, increasing Q , and a very large laser pulse occurs.

*More precisely, Q is the ratio of the total energy in the cavity to the energy lost per cycle.

[†] A dye with this property is called a saturable dye and is said to have been bleached by absorption.

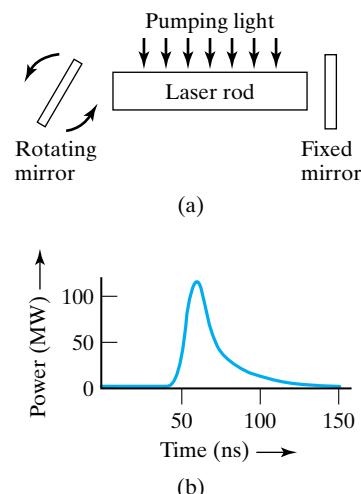
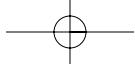


FIGURE 11.17

(a) One type of Q -switching uses a rotating mirror, which prevents laser action until the instant the two mirrors are parallel.

(b) Oscilloscope trace showing output power against time for a laser of this type.



360 Chapter 11 • Atomic Transitions and Radiation

Coherence Length

To conclude this section, we return to the property of coherence of laser light as compared to light from other sources. In particular, we introduce the notion of **coherence length** as a measure of this property. To better understand the notion of coherence, let us consider first a perfectly sinusoidal wave. Such a wave has the property that its oscillations at any one point A are perfectly correlated, or coherent, with the oscillations at any other point B no matter how far A is from B . This means, for example, that if the distance $AB = n\lambda$ (with n an integer, small or large), the oscillations at A and B are exactly in step; and if we could somehow combine the waves from A and B , we would obtain perfectly constructive interference. Similarly, if $AB = (n + \frac{1}{2})\lambda$, the oscillations at A and B are exactly out of step and, if combined, would interfere perfectly destructively.

No real wave is perfectly sinusoidal. Among other things, it is found that light waves from most sources are produced with continual random shifts in their phase. (For instance, in an ordinary light filament, photons emitted in different regions are totally unrelated and have random relative phase.) This means that the relative phase of the oscillations at any two points A and B is continually and randomly changing. As one might expect, these random fluctuations are small if A and B are close together, but become steadily larger as the distance AB increases. Thus if $AB = n\lambda$ with n equal to a *small* integer, the oscillations at A and B remain very nearly in step and, if combined, would interfere almost completely constructively. On the other hand, if AB is sufficiently *large*, the relative phase of the oscillations at A and B fluctuates by 180° or more. In this case, even if AB is exactly $n\lambda$, the oscillations at A and B will sometimes be in step and sometimes completely out of step. If combined, these waves would interfere, sometimes constructively and sometimes destructively. Since the fluctuations are rapid, the observed intensity would be the average of the maximum and minimum intensities (namely, half the maximum). Evidently, once the distance AB is large enough, we would get this same result whether or not $AB = n\lambda$. Under these conditions we say that the oscillations at A and B are no longer coherent. Roughly speaking, we are going to define the coherence length of a wave as the largest distance AB for which the oscillations at A and B are still coherent.

We can test these ideas using a Michelson interferometer, as sketched in Fig. 11.18. The half-silvered mirror splits the beam into two parts, which travel out and back along arms of lengths l_1 and l_2 , and are then recombined and measured by a photocell. Since the two waves started as a single wave and traveled

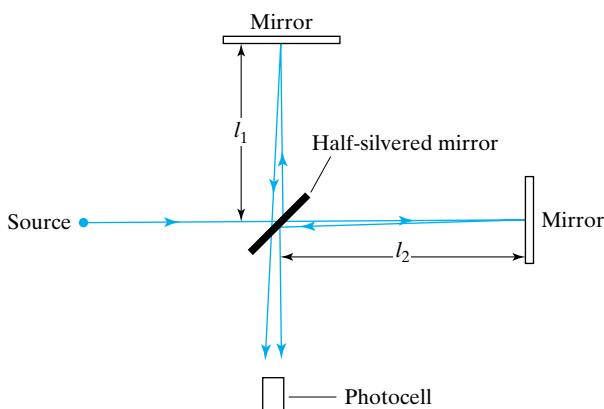
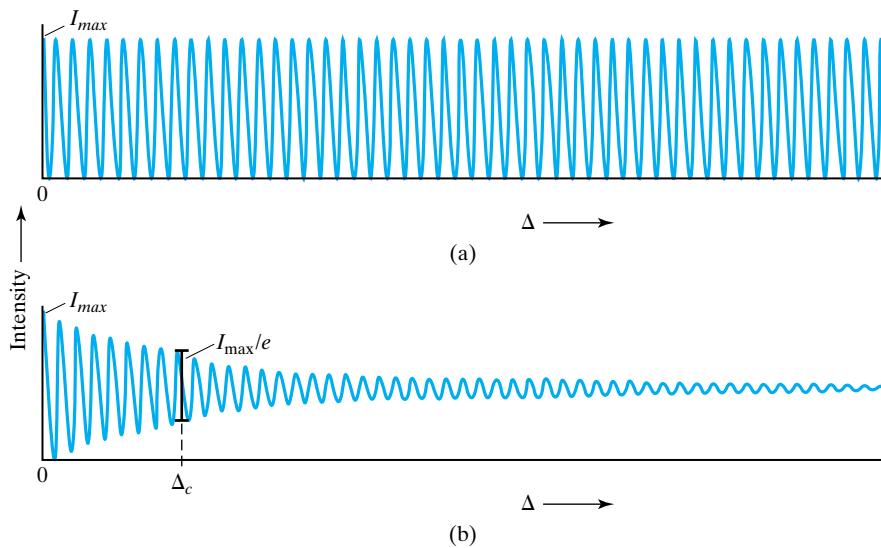
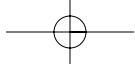


FIGURE 11.18

A Michelson interferometer. Light is split into two beams, which travel different distances, $2l_1$, and $2l_2$, before recombining.

**FIGURE 11.19**

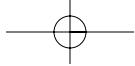
Intensity I as a function of path difference Δ in a Michelson interferometer. (a) If the light source were perfectly coherent, I would alternate between I_{\max} and back to I_{\max} each time Δ increased by one wavelength. (b) For any real source, one finds that as Δ increases, the coherence diminishes and the difference between the maxima and minima slowly decreases. The coherence length Δ_c can be defined as the value of Δ for which this difference has decreased to I_{\max}/e .

distances that differ by $\Delta = 2(l_1 - l_2)$, the photocell is in effect measuring the superposition of the oscillations at two points, A and B , in the same wave, the distance between A and B being the path difference Δ . By varying the length of either arm, we can study the interference as a function of the distance $AB = \Delta$.

If the original wave were perfectly sinusoidal, then as Δ increases, the observed intensity would alternate between maxima, all of the same height I_{\max} , and minima, $I_{\min} = 0$, as in Fig. 11.19(a). These alternations would continue no matter how large we made Δ . In reality, the wave is not perfectly sinusoidal and the situation is as shown in Fig. 11.19(b). When Δ is small, the intensity alternates between I_{\max} and 0, much as in Fig. 11.19(a). But as Δ increases, the two waves begin to lose their coherence, and the maxima begin to shrink and the minima start to grow. Finally, as $\Delta \rightarrow \infty$, the contrast between the maxima and minima disappears entirely, and I approaches the constant value $I_{\max}/2$.

It is clear from Fig. 11.19(b) that one cannot define a unique coherence length Δ_c , beyond which the two waves abruptly lose their coherence. Rather, the coherence fades out continuously, and one can, for example, define Δ_c as the distance at which the difference between successive maxima and minima has dropped to the value I_{\max}/e . This is the definition indicated in Fig. 11.19(b).

Before the advent of lasers, the coherence length of typical light sources was less than a millimeter or so, although lengths of order 10 cm could be achieved with difficulty. Since the coherence length of lasers can easily exceed 1 km, they have enormously increased the usefulness of interferometers for measuring distances. In a typical application one arm of the interferometer is kept fixed while the mirror on the other is moved through the distance that is to be measured. In this way, distances can be measured to a fraction of a wavelength. However, since interference cannot be observed once Δ is much more than Δ_c , the distances that can be measured cannot be much larger than Δ_c . Therefore, the laser has increased by many orders of magnitude the distances that can be measured using interferometers. Perhaps equally important, laser interferometers can measure displacements of objects that are far removed from the interferometer (and hence have $l_2 \gg l_1$), as with the arrangement shown in Fig. 11.20.



362 Chapter 11 • Atomic Transitions and Radiation



FIGURE 11.20

In this laser interferometer one of the paths is inside the device while the other extends to a mirror attached to the point whose displacement is to be measured. This photo shows a reading of 1 microinch, the displacement of the I-beam caused by the weight of a penny.

CHECKLIST FOR CHAPTER 11

CONCEPT	DETAILS
Radiation from a classical charge	$P = 2kq^2a^2/3c^3$ (11.1)
Inevitable collapse of classical atom	Lifetime $\sim 10^{-11}$ s (11.4)
Absence of radiation from stationary states	Static charge distribution (Sec. 11.3)
Transitions caused by perturbations	Upward = absorption of a photon Downward = emission of a photon
Hamiltonian operator	$H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x)$ (11.11) (with suitable generalizations)
Time-independent Schrödinger equation	$H\psi = E\psi$ (11.13)
Time-dependent Schrödinger equation	$i\hbar\partial\Psi/\partial t = H\Psi$ (11.16)
Completeness of the stationary-state wave functions	Any $\psi(x) = \sum A_n\psi_n(x)$ (11.20) $ A_n ^2$ = probability of finding system in state ψ_n $A_m = \int \psi_m^*\psi dx$ (11.24)
Orthogonality	$\int \psi_m^*\psi_n dx = 0$ for $m \neq n$ (11.23)
Normalization	$\int \psi_n^*\psi_n dx = 1$
Time-dependent perturbation theory	Theory of transitions due to a perturbation W (Sec. 11.5)
Transition probability	$P(n \rightarrow m) \propto \int \psi_m^* W \psi_n dx ^2$ (11.40) and (11.43)
Stimulated emission	Emission of a photon in a downward transition stimulated by applied radiation
Spontaneous emission	Emission of a photon in a downward transition of an isolated atom “stimulated” by the vacuum fluctuation fields
Selection rule	Any general rule specifying forbidden transitions (Table 11.1)

Metastable state	An excited state with no allowed downward transitions (Sec. 11.8)
Lasers	Pulsed and continuous (Sec. 11.9) Angular divergence, time dependence, Q -switching, coherence length (Sec. 11.10)

PROBLEMS FOR CHAPTER 11

SECTION 11.2 (Radiation by Classical Charges)

- 11.1** • A charge q executes simple harmonic motion with position $x = x_0 \sin \omega t$. **(a)** Find P , the total power of the radiation emitted by this oscillating charge. **(b)** Show that the average power over one complete cycle is

$$P_{av} = \frac{kq^2\omega^4x_0^2}{3c^3}$$

- 11.2** • In the antenna of a TV or radio station charges oscillate at some frequency f and radiate electromagnetic waves of the same frequency. As a simple model of such an antenna, imagine that a single charge $q = 250 \text{ nC}$ is executing simple harmonic motion at 100 MHz with amplitude 0.3 m. ($1 \text{ nC} = 10^{-9} \text{ coulomb}$.) Use the result of Problem 15.1 to calculate the average total power radiated by this antenna.

- 11.3** • A certain cell phone transmitter radiates 3 W of power at about 900 MHz. **(a)** Find the rate of emission of photons by the transmitter. **(b)** In this case is there an appreciable difference between the correct quantum view and the classical picture in which the radiation is emitted continuously?

- 11.4** • Carry out the calculation of Example 11.1 using SI units throughout. [Note: If you use a calculator and exceed its maximum exponent range (usually 10^{+99}), you should calculate the mantissa and exponent separately.]

- 11.5** •• The formula (11.1) for the power P radiated by an accelerating charge q can be derived by the method of **dimensional analysis**. Since P would be expected to involve k , q , a , and c , we might reasonably guess that it should have the form

$$P = bk^l q^m a^n c^p \quad (11.48)$$

where b is some dimensionless number of order 1 (perhaps something like 4π) and where l, m, n , and p are unknown powers (and k is the Coulomb force constant $\approx 9 \times 10^9 \text{ N} \cdot \text{m}^2 \cdot \text{C}^{-2}$). By inserting their units into the five dimensional quantities in (11.48), you will get an equation that determines the unknown powers l, m, n , and p . Show that you obtain the correct form (11.1), except that the dimensionless number b cannot be determined.

- 11.6** •• The result (11.3) shows that a classical atom would collapse very rapidly. To get a very rough estimate of the time for a hydrogen atom to collapse completely, do the following: **(a)** You can find the rate at which the radius r shrinks,

$$\frac{dr}{dt} = \frac{dr}{dE} \frac{dE}{dt} \quad (11.49)$$

with dr/dE determined from (5.10) and $dE/dt = -P$. Find dr/dt when $r = a_B$. **(b)** Making the admittedly crude approximation that that dr/dt remains constant, estimate roughly how long the electron takes to spiral in from $r = a_B$ to $r = 0$. (For a more realistic estimate, see Problem 11.15.)

- 11.7** •• Many particle accelerators, including the cyclotron and the synchrotron (Sections 17.11 and 18.11), hold charged particles in a circular orbit using a suitable magnetic field. The centripetal acceleration, $a = v^2/r$, can be very large and can lead to serious energy loss by radiation, in accordance with Eq. (11.1). **(a)** Consider a 10-MeV proton in a cyclotron of radius 0.5 m. Use the formula (11.1) to calculate the rate of energy loss in eV/s due to radiation. **(b)** Suppose that we tried to produce electrons with the same kinetic energy in a circular machine of the same radius. In this case the motion would be relativistic and formula (11.1) is modified by an extra factor* of γ^4 :

$$P = \frac{2kq^2a^2\gamma^4}{3c^3} \quad (11.50)$$

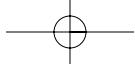
Find the rate of energy loss of the electron, and compare with that for a proton. (Your answer for the electron should be enormously larger than for the proton. This explains why most electron accelerators are linear, not circular, since the acceleration in a linear accelerator — once $v \sim c$ — is far smaller than the centripetal acceleration considered here.)

- 11.8** •• Answer the same questions as in Problem 11.7, but assume that both the proton and electron have kinetic energy 10 GeV and move in a circle of radius 20 m. [In this case both particles are relativistic and you must use the relativistic formula (11.50).]

- 11.9** •• The ring called PEP-II at Stanford in California stores electrons orbiting around a circle† of radius 170 m. Because of the centripetal acceleration, $a = v^2/r$, the particles lose energy in accordance with Eq. (11.50) [which is the appropriate relativistic form of Eq. (11.1)]. **(a)** Find the rate of energy loss of a single 9-GeV electron. **(b)** If a total of 2×10^{12} electrons are radiating at this rate, what is the total power, in watts, needed to keep them at 9 GeV? (For comparison, the power used by a typical household appliance is on the order of 100 W.)

*Note that this is for the case of circular motion. For linear motion the factor is γ^6 .

†The actual device contains both curved and straight sections but is reasonably described as a single circle for the purposes of this problem.



364 Chapter 11 • Atomic Transitions and Radiation

11.10 •• Figure 11.21 shows an electric field line from a charge q that was kicked abruptly from rest and has moved a distance d . The kink in the field line occurs between r and $r + \Delta r$, where $r \approx ct$ and t is the time since q was kicked. The radiation from q is carried by the transverse component of \mathcal{E} , shown as \mathcal{E}_{tr} in the picture. Prove that as a function of θ , \mathcal{E}_{tr} is proportional to $\sin \theta$. (This means that the maximum radiation is at $\theta = 90^\circ$ and there is no radiation at $\theta = 0$. For this reason, broadcasting antennas are oriented at 90° to the direction in which they need to transmit.)

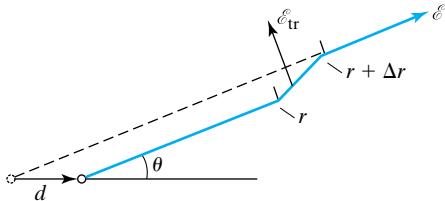


FIGURE 11.21

(Problem 11.10).

11.11 •• (a) A classical point charge q of mass m is in a circular orbit of radius r around a fixed charge Q (with q and Q of opposite sign, of course). Starting from Eq. (11.1), derive a formula for the radiated power P in terms of q , m , r , and Q . **(b)** By what factor is P changed if we double q (leaving m , r , and Q unchanged)? **(c)** What if we double Q (with m , r , and q unchanged)?

11.12 •• One of the difficulties with classical models of the atom was that they failed to predict the correct frequency for the radiation emitted. According to classical electromagnetic theory, the frequency of emitted radiation should equal the frequency of the orbiting electron. **(a)** Calculate the orbital frequencies, $f_{\text{orb}}(1)$ and $f_{\text{orb}}(2)$, of a classical electron in the $n = 1$ and $n = 2$ Bohr orbits of a hydrogen atom. **(b)** Now find the frequency $f_\gamma(2 \rightarrow 1)$ of the actual photon emitted in the $2 \rightarrow 1$ transition. Show that $f_\gamma(2 \rightarrow 1)$ is *not* equal to either $f_{\text{orb}}(2)$ or $f_{\text{orb}}(1)$ (or their average or their difference). **(c)** It turns out, however, that as $n \rightarrow \infty$ the orbital frequency $f_{\text{orb}}(n)$ of the n th orbit does approximate the frequency $f_\gamma(n \rightarrow n - 1)$ of a photon emitted in the transition $(n \rightarrow n - 1)$. Prove that $f_{\text{orb}}(n) = E_R / (\pi \hbar n^3)$. Derive an expression for $f_\gamma(n \rightarrow n - 1)$, and show that it approaches $f_{\text{orb}}(n)$ as $n \rightarrow \infty$. (This result — called the **correspondence principle** — played an important role in the development of the Bohr model.)

11.13 •• Consider the electron in a classical He^+ ion. Using the method of Example 11.1, find the radiated power predicted by the classical radiation formula (11.1) for a radius equal to that of the first Bohr orbit of He^+ . Compare your result with that predicted for hydrogen.

11.14 ••• Using the method of Problem 11.6, estimate the lifetime of a classical He^+ ion starting in the $n = 1$ Bohr orbit for He. Compare with the answer for hydrogen.

11.15 ••• Problem 11.6 estimates the time Δt for collapse of a classical hydrogen atom, making the approximation that it shrinks at a constant rate. Calculate Δt

without making that approximation, as follows: **(a)** Use Eq. (11.49) to find dr/dt , the rate at which the orbit shrinks, as a function of r , and find the time for this classical atom to collapse entirely, by evaluating

$$\Delta t = \int_{a_B}^0 \frac{dr}{dr/dt}$$

SECTION 11.4 (More Quantum Formalism*)

11.16 • The Hamiltonian operator is often described as a **linear operator** because it has the linear property that $H[A\psi(x) + B\phi(x)] = AH\psi(x) + BH\phi(x)$ for any two functions $\psi(x)$ and $\phi(x)$ and any two constant numbers A and B . Prove that the one-dimensional Hamiltonian (11.11) is linear.

11.17 •• Show that the wave functions (11.21) for the infinite square well satisfy the orthogonality property that $\int_0^a \psi_m(x)^* \psi_n(x) dx = 0$ for $m \neq n$.

SECTION 11.5 (Transitions; Time-Dependent Perturbation Theory*)

11.18 • (a) The first excited state of the sodium atom is 2.11 eV above the ground state. What wavelength radiation can cause transitions between these two levels? What sort of radiation is this? (visible, UV, etc.?) **(b)** Answer the same questions for the two $2p$ levels in hydrogen, which are shown in Fig. 9.7 and are 4.5×10^{-5} eV apart. (This is the fine-structure splitting discussed in Section 9.7.) **(c)** Answer the same questions for the lowest two levels of the ${}^7\text{Li}$ nucleus, which are 0.48 MeV apart.

11.19 • The atoms of a certain monatomic gas have five energy levels: $E_1 = 0$, $E_2 = 5.4$, $E_3 = 8.2$, $E_4 = 8.6$, and $E_5 = 12.4$, all measured up from E_1 in eV. **(a)** If infrared light with wavelengths in the range from 3000 to 3200 nm shines through the gas, what transitions could it cause? If the gas was so cool that all atoms were in the ground state, would you expect to observe these transitions? **(b)** Answer the same questions for UV with wavelengths from 95 to 105 nm. **(c)** What wavelength light could de-excite the atoms from level 3 to level 2?

11.20 • The wave functions for an infinite square well, with edges at $x = 0$ and $x = a$, are given in (11.21). Prove that if we move the origin to the middle of the well, these same functions are as given in (11.34) and (11.35). (Actually, some of the new functions differ by an unimportant sign from the original ones.)

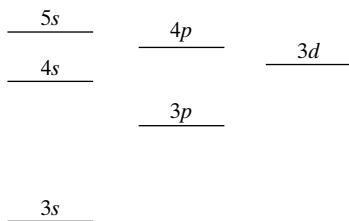
11.21 •• Confirm the result (11.37) for the probability $P(1 \rightarrow 2)$ for a transition from level 1 to level 2 of the rigid box, by evaluating the integral (11.36).

11.22 •• Consider an electron that is initially in the first excited state of the infinite square well of Example 11.4 (Sec. 11.5). **(a)** If an electric field \mathcal{E} is switched on in the x direction, find the probability $P(2 \rightarrow 1)$ that the electron will be found in the ground state a short time Δt later. **(b)** Similarly, find the probability $P(2 \rightarrow 3)$ for excitation to the second excited state. **(c)** What are $P(2 \rightarrow 4)$ and $P(2 \rightarrow 5)$? Compare these various probabilities.

SECTION 11.8 (Atomic Selection Rules)

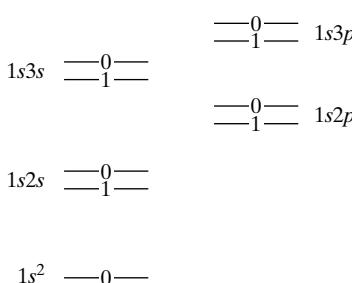
11.23 • When a quantum wave function Ψ is complex (with both real and imaginary parts), its probability density is $|\Psi|^2$, where $|\Psi|$ is the absolute value of Ψ , defined by Eq. (6.12) as $|\Psi| = \sqrt{\Psi_{\text{real}}^2 + \Psi_{\text{imag}}^2}$. Prove that $|\Psi|^2 = \Psi^* \Psi$, where Ψ^* is the complex conjugate of Ψ . (The complex conjugate z^* of any complex number $z = x + iy$, where x and y are real, is defined as $z^* = x - iy$.)

11.24 • The outermost (valence) electron of sodium is in a $3s$ state when the atom is in its ground state (Table 10.2 in Section 10.7). The valence electron can be excited to higher levels, the first few of which are shown in Fig. 11.22. Given the selection rule that only those transitions for which $\Delta l = \pm 1$ are allowed, indicate on this energy-level diagram all allowed transitions among the levels shown.

**FIGURE 11.22**

Some low-lying levels of any one electron in sodium (Problem 11.24).

11.25 • Figure 11.23 shows some of the lowest energy levels of the He atom. They are labeled by their configuration (for example, $1s2p$ means that the atom has one electron in the $1s$ level and one in the $2p$ level). The energy depends somewhat on the orientation of the two electrons' spins: If the spins are antiparallel, the total spin is zero (quantum number $s_{\text{tot}} = 0$); if the spins are parallel, the total spin has $s_{\text{tot}} = 1$. For a given configuration, the state with $s_{\text{tot}} = 1$ has slightly lower energy. **(a)** Explain why the $1s^2$ configuration has only $s_{\text{tot}} = 0$. **(b)** There is a selection rule $\Delta s_{\text{tot}} = 0$, that is, transitions in which s_{tot} changes are forbidden. Indicate all allowed transitions on the energy-level diagram. (Don't forget the selection rule $\Delta l = \pm 1$.) **(c)** Which excited levels would you expect to be metastable?

**FIGURE 11.23**

Some low levels of helium. (These are the energy levels of the whole atom, not the individual electrons.) The numbers 0 and 1 indicate the quantum number s_{tot} of the total spin (Problem 11.25).

11.26 • The general proof of the selection rule (11.46) is beyond our scope, but you can prove it in a few special cases. **(a)** It is a fact that the wave function $\psi_{nlm}(\mathbf{r}) = R_{nl}(r)\Theta_{lm}(\theta)e^{im\phi}$ satisfies

$$\psi_{nlm}(-\mathbf{r}) = (-1)^l \psi_{nlm}(\mathbf{r}) \quad (11.51)$$

[This property is often described by saying that the wave function has **parity** $(-1)^l$.] Use the angular functions listed in Table 8.1 to prove (11.51) for all wave functions with $l = 0, 1$, or 2 . **(b)** The probability of a radiative transition $(n, l, m \rightarrow n', l', m')$ is given by (11.45), which now takes the form

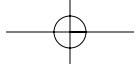
$$P(n, l, m \rightarrow n', l', m') \propto \mathcal{E}_0^2 \left| \int \psi_{n'l'm'}^* x \psi_{nlm} dV \right|^2 \quad (11.52)$$

(This is for radiation polarized with \mathcal{E} in the x direction. For isotropic unpolarized radiation, we must average over this and the two corresponding expressions with x replaced by y and by z .) Use the property (11.51) to prove that the transition probability (11.52) is zero if $l' = l$ (whether the integrand is x, y , or z). This proves that transitions with $\Delta l = 0$ are forbidden — a particular case of (11.46).

11.27 • There is a selection rule on the total angular momentum (all spins plus all orbital momenta) of an atom or nucleus. The total angular momentum is given by a quantum number j_{tot} , and it is found that transitions which do not satisfy $\Delta j_{\text{tot}} = 0$ or ± 1 are forbidden (that is, very improbable). Further, among the forbidden transitions the larger the value of Δj_{tot} , the more improbable a transition is. This trend applies to β decay in nuclei as well as to radiative transitions. Use these facts to explain why an excited state of the ${}^{180}\text{Ta}$ nucleus occurs naturally in measurable amounts. (See Appendix D.)

11.28 • Consider two energy levels of the helium atom, in both of which the two electrons' spins are antiparallel (so that the total spin is zero, and the spins can be ignored) and one of the electrons is in the lowest ($1s$) orbital. In the upper level the second electron has $l = 2$; in the lower level the second electron has $l = 1$. The atom is placed in a magnetic field, and (as described in Section 9.4) the upper level splits into five equally spaced sublevels and the lower into three sublevels (with the same spacing). **(a)** Sketch the resulting levels. **(b)** There are, in principle, 15 different possible transitions from the upper ($l = 2$) level to the lower ($l = 1$) level. Show that because the sublevels all have the same spacing, there are actually only seven distinct energy differences. **(c)** The selection rules for these transitions are $\Delta l = \pm 1$ and $\Delta m = 0$ or ± 1 ; that is, only transitions that satisfy these rules are allowed. Indicate all of the allowed transitions on your energy level diagram. **(d)** How many distinct photon frequencies will result from allowed transitions between the two levels? This is the normal Zeeman effect described in Section 9.4.

11.29 • Carry out the same tasks as in Problem 11.28 (with the numbers modified where necessary), using the same upper level ($l = 2$) but with a lower level in which one electron is in an s state and the other in an f state ($l = 3$).



366 Chapter 11 • Atomic Transitions and Radiation

11.30 •• The probability of a radiative transition ($n, l, m \rightarrow n', l', m'$) induced by unpolarized isotropic radiation is given by an average of (11.52) and the two corresponding expressions with x replaced by y and by z . Prove the selection rule that this probability is zero unless $m' = m$ or $m \pm 1$ (that is, $\Delta m = 0$ or ± 1). [Hint: Use the form (8.98) of the wave functions and write the integral of (11.52) in terms of spherical polar coordinates. You will need to examine only the integral over ϕ .]

11.31 ••• The selection rule (11.46) that radiative transitions are forbidden unless $\Delta l = \pm 1$ is related to the fact that photons have spin 1, as we mentioned in Section 11.8. The details of the argument are surprisingly complicated and require a greater knowledge of quantum mechanics than we have developed. Nevertheless, you can reach the right conclusion arguing semiclassically, as follows: To be definite, consider a hydrogen atom whose electron has initial orbital angular momentum \mathbf{L} and spin \mathbf{S} and which is excited by a photon to a state, with \mathbf{L}' and \mathbf{S}' . Conservation of angular momentum requires that

$$\mathbf{L}' + \mathbf{S}' = \mathbf{L} + \mathbf{S} + \mathbf{L}_\gamma + \mathbf{S}_\gamma \quad (11.53)$$

where \mathbf{L}_γ and \mathbf{S}_γ are the orbital and spin angular momenta of the incoming photon. Notice first that the electric field has no effect on the electron's spin,* so that $\mathbf{S}' = \mathbf{S}$ and these two terms cancel from (11.53). **(a)** Next you can show that $\mathbf{L}_\gamma = 0$, as follows: If it is to interact with the atom, the incoming photon must come within a distance of order a_B of the nucleus; thus, to interact with the atom, its orbital angular momentum must be of order $a_B p$ or less, where p is the linear momentum of the photon. Given that the photon's energy has to be of order E_R , show that L_γ must be of order $\alpha\hbar$ or less, where α is the so-called fine-structure constant $\alpha = ke^2/\hbar c \approx 1/137$. Since the only possible values of L are $\sqrt{l(l+1)}\hbar$ with l equal to an integer, this shows that $l = 0$ and $L_\gamma = 0$. Thus, (11.53) reduces to

$$\mathbf{L}' = \mathbf{L} + \mathbf{S}_\gamma \quad (11.54)$$

(b) The maximum magnitude of \mathbf{L}' occurs when \mathbf{L} and \mathbf{S}_γ are parallel. By taking the z component of (11.54) and remembering that the maximum value of L_z is $l\hbar$ (and similarly for all angular momenta), show that the maximum possible value of l' is $l + 1$. Show similarly that the minimum possible value is $l - 1$. This leaves only three possibilities, $l' = l + 1$ or l or $l - 1$, and Problem 11.26 has shown that $l' = l$ is not allowed, so you're home.

SECTION 11.10 (Further Properties of Lasers*)

11.32 • A single-mode He–Ne laser has a beam of diameter 3 mm. As the beam propagates away from the laser, its diameter increases because of diffraction. **(a)** Given that this effect simply adds to the initial beam size, find the distance at which the beam diameter is doubled. **(b)** At what distance will the beam have a diameter of 1 m?

*Recall that the spin enters the energy through the term $-\boldsymbol{\mu} \cdot \mathbf{B}$, which involves the magnetic field, and that the effects of B are usually negligible compared to those of \mathcal{E} .

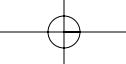
11.33 •• In three of the Apollo lunar experiments, astronauts left reflector panels on the moon so that laser beams from the earth could be reflected off the panels and back to earth. Lasers with $\lambda = 532$ nm send pulses of 0.3 J to the moon, and the round-trip time is measured within $\delta t \approx 0.4$ ns. In this way the one-way distance to the reflector on the moon is determined regularly within $\delta l = c\delta t/2 \approx 6$ cm. The beams from the lasers have a diameter of 5 km on the moon. (This is the result of spreading due to diffraction and to atmospheric turbulence.) The reflector panels contain 300 mirrors, each of diameter 4 cm. **(a)** Find the number of photons sent from the earth in a single pulse. **(b)** Find the fraction of these photons that strike any one of the small mirrors. **(c)** Find the angular divergence $\delta\theta \approx \lambda/d$ of the reflected beam from this mirror (where d is the mirror's diameter). **(d)** What is the diameter of the reflected beam when it returns to earth? **(e)** The return light is measured by a photomultiplier at the focus of a telescope. What fraction of the return light is captured by the telescope, whose diameter is 1 m? **(f)** For a single pulse, find the total number of photons captured by the telescope from all 300 mirrors. (The actual number is somewhat smaller because of losses in the atmosphere and in the telescope.) This experiment monitors the moon's orbit within a few centimeters (compared to the earth–moon distance of 4×10^8 m), allowing stringent tests of competing theories of gravity.

COMPUTER PROBLEMS

11.34 •• (Section 11.4) To illustrate the property (11.20) of completeness, consider an infinite square well of width $a = 1$, centered on the origin, with stationary-state wave functions $\psi_n(x)$ as given in (11.34) and (11.35). Let $\psi(x)$ be the flat wave function equal to 1 everywhere inside the well (and 0 outside). **(a)** Find the coefficients A_m in the expansion (11.20) of $\psi(x)$. **(b)** To reproduce $\psi(x)$ exactly, you would have to include all of the terms in the infinite series (11.20), but you can get a surprisingly good approximation by including just the first few terms. Use a suitable plotting program to plot the sum of the first 5 nonzero terms of the series. **(c)** Repeat with the first 10 nonzero terms.

11.35 •• (Section 11.4) Do the same exercises as in Problem 11.34 for the same square well with width 1, but using the triangular wave $\psi(x) = \sqrt{2}|(1 - 2x)|$. (Sketch this wave function first.)

11.36 ••• (Section 11.5) Consider the electron of Problem 11.22, which is initially in the $n = 2$ state ψ_2 of the infinite square well. When an electric field \mathcal{E} is switched on for a brief time Δt , the wave function picks up a small component proportional to $x\psi_2(x)$. **(a)** Use a suitable plotting program to make plots of $\psi_2(x)$ and $x\psi_2(x)$. **(b)** It is probably not immediately clear what the extra wave function $x\psi_2(x)$ (multiplied by a small constant) signifies. To clarify this, expand $x\psi_2(x)$ as in (11.20), and find the coefficients A_m of this expansion. Use your answer to argue that, to a good approximation, the extra wave function has the form $x\psi_2(x) \approx A_1\psi_1(x) + A_3\psi_3(x)$. Plot the right-hand side of this approximation, and compare with your plot of $x\psi_2(x)$. What is the physical significance of this result?



SYSTEMS WITH TWO OR MORE ATOMS



- Chapter 12 Molecules
- Chapter 13 Solids — Theory
- Chapter 14 Solids — Applications
- Chapter 15 Statistical Mechanics

In Part III (Chapters 12 to 15) we treat applications of quantum mechanics to systems that are larger than a single atom. In Chapter 12 we discuss how atoms combine to form molecules. Traditionally, this topic is one of the principal concerns of chemistry, but the boundary between physics and chemistry is now quite blurred. Whether we regard the study of molecules as part of physics or chemistry, it is certainly an important application of quantum mechanics.

In Chapters 13 and 14 we describe how atoms or molecules combine to form solids. The solid is our first example of a macroscopic system for whose understanding we require quantum mechanics. Solid-state physics has spawned a dazzling array of electronic “solid-state” devices — transistors, LEDs, photovoltaic cells, integrated circuits — that have made possible the electronic revolution of the last few decades. Chapter 13 is mainly about the theory of solids, and Chapter 14 about some of these applications.

Finally, Chapter 15 is an introduction to statistical mechanics, the statistical treatment of systems such as gases and solids, usually containing very large numbers of atoms or molecules.

As far as possible, the chapters of Parts III and IV are designed to be independent of one another. For example, you could jump directly to Chapter 15 without reading Chapters 12 through 14. If you want to read the chapters on solids (13 and 14), you would need to read just the first four sections of Chapter 12. If you are interested only in the foundations of solid-state physics, and not in the technological applications, you could read Chapter 13 only and skip 14. However, if you wish to understand the applications, you should first read at least the first 5 sections of Chapter 13 before proceeding to Chapter 14. If your main interest is in nuclear or particle physics, you could skip Part III entirely and proceed straight to Part IV.

PART III

Chapter 12

Molecules

- 12.1** Introduction
 - 12.2** Overview of Molecular Properties
 - 12.3** The Ionic Bond
 - 12.4** The Covalent Bond
 - 12.5** Directional Properties of Covalent Bonds*
 - 12.6** Excited States of Molecules*
 - 12.7** Molecular Spectra*
- Problems for Chapter 12

*These sections can be omitted without significant loss of continuity

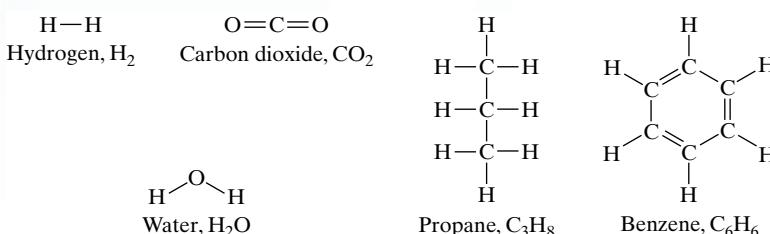
12.1 Introduction

A molecule is a stable, or nearly stable, bound state of two or more atoms. An oxygen molecule, O_2 , is a bound state of two oxygen atoms; a water molecule, H_2O , is a bound state of two hydrogens and one oxygen. In this chapter we discuss the nature of the forces that bind atoms to one another and some of the properties of the resulting molecules.

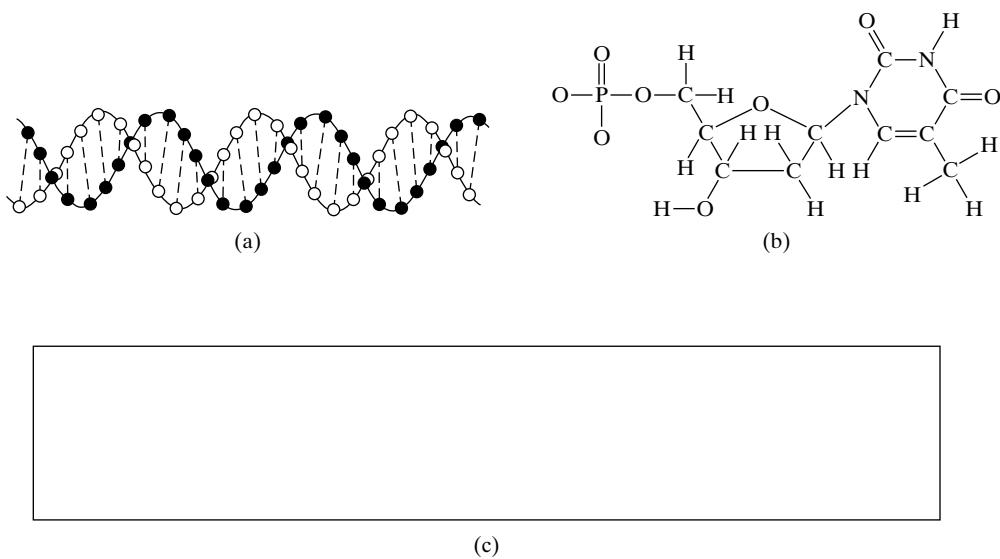
The number of atoms in a molecule ranges from two, in a **diatomic molecule** such as O_2 or CO, to several billions in large biological molecules. In Fig. 12.1 we have illustrated the arrangement of the atoms in five different, relatively simple molecules, and Fig. 12.2 shows one of the largest known molecules, DNA, which carries genetic information in organisms. Even larger numbers of atoms can form stable bound states. For example, if we took any large number of sodium atoms and the same number of chlorine atoms, then at room temperature these could bind together to form a stable salt crystal; but we would usually describe such a crystal as a *solid*, not a *molecule*.* If, however,

FIGURE 12.1

The atoms within molecules are arranged in many different ways: straight lines, triangles, chains, rings, and many others. The lines connecting atoms represent molecular bonds, and double lines indicate double bonds, as described in Section 12.4.



*There is no clear unambiguous distinction between molecules and solids. In fact, one could argue that a salt crystal is just an enormous molecule. Nonetheless, this is not normal usage. In particular, it is generally considered that any given molecular species should have a definite number of constituent atoms, whereas one can add any number of Na–Cl pairs to a salt crystal and still have a salt crystal which is just a little larger. By this criterion, then, a salt crystal is not a molecule.

**FIGURE 12.2**

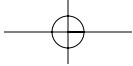
(a) A segment, about 10 nm long, of the double helix of the DNA molecule. **(b)** Each circle in (a) represents a base, comprising about 30 atoms, and there are altogether just four different kinds of bases. This picture shows the atomic composition of the base called thymine. **(c)** Electron micrograph of a segment of DNA about 10 μm long (coated in platinum to make it show up). Although this segment is about a thousand times longer than the segment shown in (a), it is still only a tiny fraction of an entire molecule. The average DNA molecule in a human chromosome is about 4 cm long and contains some 10 billion atoms.

we heated the salt until it vaporized, we would find that the atoms would move apart in bound pairs, one Na atom bound to each Cl, and these pairs are exactly what we do describe as molecules of NaCl. In this chapter we discuss only molecules, such as the NaCl molecule, containing a definite and reasonably small number of atoms. Further, we will consider only individual, isolated molecules. Since the molecules in a solid or liquid can never really be considered to be isolated, this means that this chapter is actually about molecules in a gas. The properties of solids and (very briefly) liquids are discussed in Chapters 13 and 14.

A feature common to all molecules is that the force which holds their atoms together is ultimately the electrostatic attraction between opposite charges. At first, one might think there could be no electrostatic force between two neutral atoms. However, when two or more atoms are sufficiently close, their charges can redistribute themselves with pairs of opposite charges close together, so that attraction predominates over repulsion, bonding the atoms together to form a molecule.

Although all interatomic bonds are basically electrostatic, we can distinguish five different ways in which the necessary redistribution of charge occurs. The resulting kinds of bond are called:

- the ionic bond
- the covalent bond
- the hydrogen bond
- the van der Waals bond
- the metallic bond



370 Chapter 12 • Molecules

Gerhard Herzberg
(1904–1999, German-Canadian)



One of the giants of molecular physics, Herzberg — a chemist — is perhaps best known for his definitive four-volume work on molecular spectra and structure. Born in Hamburg and educated at Göttingen and Berlin, he fled Nazi Germany in 1935 and moved to Canada. His painstaking analyses of molecular spectra made possible numerous applications, such as the identification of many molecules in interstellar space. He won the Nobel Prize in chemistry in 1971.

Ionic and covalent bonds are important in simple molecules and are the main topics of this chapter. The hydrogen bond involves the sharing of a proton between two negative ions. Since its main role is in large organic molecules, which are beyond the scope of this book, we will not discuss the hydrogen bond here. The van der Waals bond is very weak and is usually important only when other types of bond do not act; it is the bond that holds some liquids and solids together and is discussed briefly in Chapter 13. As its name implies, the metallic bond is what binds metals together and is described in Chapter 13.

In this chapter we focus mostly on diatomic molecules since they are relatively simple but still exemplify many of the general principles of molecular binding. We begin with a brief overview of molecular properties in Section 12.2. In Sections 12.3 and 12.4 we give more detailed descriptions of the ionic and covalent bonds. In Section 12.5 we discuss how some molecular bonds have pronounced directional properties. This gives many molecules distinct shapes, which can be predicted using the properties of atomic wave functions described in Chapters 8 and 10. In Section 12.6 we discuss the excited states of molecules. We will see that the distribution of the energy levels in a molecule is quite different from that in an atom. For this reason, molecular spectra are quite different from typical atomic spectra, as we describe in Section 12.7.

12.2 Overview of Molecular Properties

The attraction that binds atoms together to form molecules is the electrostatic attraction between the positive nucleus of each atom and the negative electrons in all the other atoms. Because ordinary matter is electrically neutral, we are usually unaware of the enormous magnitude of the internal electric forces. It is only when the positive and negative charges are rearranged into separate regions that electric forces become evident. Such rearrangement is what happens within molecules on the microscopic scale. To illustrate the magnitude of the resulting forces, we first consider a thought experiment on a macroscopic scale.

Example 12.1

Imagine the charges in 1 gram of hydrogen separated so that all of the electrons are at one place and all of the protons at another, separated from the electrons by a distance equal to the earth's diameter. What would be the force of attraction between the electrons and the protons?

Each atom of hydrogen contains one electron and one proton, and 1 gram of hydrogen contains Avogadro's number, N_A , of atoms. Thus, the magnitude of the charge of either sign is

$$Q = N_A e = (6.02 \times 10^{23}) \times (1.60 \times 10^{-19} \text{ C}) = 9.63 \times 10^4 \text{ C}$$

The enormity of these charges becomes apparent when we calculate the attraction between them when separated by an earth diameter:

$$F = k \frac{Q^2}{d^2} = \left(8.99 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \right) \times \frac{(9.63 \times 10^4 \text{ C})^2}{(1.27 \times 10^7 \text{ m})^2} = 5.2 \times 10^5 \text{ N}$$

This is a force of roughly 60 tons!

TABLE 12.1

Sampling of eight diatomic molecules, four of which are bound ionically and four covalently. (Many molecules are bound by a mixture of both kinds of bonds.) The second column gives the bond length R_0 , and the third gives the binding energy (or dissociation energy) B .

Molecule	R_0 (nm)	B (eV)	Bond
KCl	0.27	4.1	Ionic
LiF	0.16	5.9	
NaBr	0.25	3.7	
NaCl	0.24	4.2	
H ₂	0.074	4.5	Covalent
HCl	0.13	4.4	
N ₂	0.11	9.8	
O ₂	0.12	5.1	

It is the strong electrostatic interaction between the many charges within ordinary matter that is responsible for the strength of solids, the contact force when one solid presses against another, and the energy released in chemical reactions such as combustion. Most important, for our present purposes, it is the electrostatic interaction that binds the atoms together within a molecule.

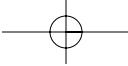
Before we try to calculate the binding strengths of molecules, let us consider some of the observed parameters for a few simple molecules. In Table 12.1 we give a sampling of eight diatomic molecules. The first column shows the chemical symbol and the second the **bond length** R_0 , defined as the distance between the two nuclei. The third column gives the **dissociation energy**, or **binding energy**, B , the energy needed to separate the molecule into two neutral atoms, and the last column gives the type of bond.

We see from Table 12.1 that typical bond lengths in diatomic molecules are one or two tenths of a nanometer, with ionic molecules tending to be a little larger than covalent. The binding energies are all of order a few eV. To understand the values of these parameters, we must examine the distribution of charge inside the molecule. This distribution depends on the behavior of the atomic electrons as the atoms approach one another, and this behavior is determined mainly by the atoms' positions in the periodic table. We now sketch three types of behavior: first, the almost complete failure of the noble gases to form bonds; second, the formation of ionic bonds between atoms at opposite sides of the periodic table; and finally, the formation of covalent bonds, often between identical or similar atoms.

The Noble Gases

Until recently, it was believed that the elements He, Ne, Ar, ... in group VIII of the periodic table formed no stable molecules, and because of this reluctance to mingle, they were named “inert” or “noble.” Although we now know that the noble atoms do form some molecules, all such molecules are very weakly bound, and the names are still perfectly appropriate.

To see why the noble atoms interact so weakly with other atoms, recall from Section 10.7 that they are closed-shell atoms and are therefore spherically symmetric. Gauss’s law for the electric field tells us that a spherically symmetric charge distribution of zero total charge produces no external electric



372 Chapter 12 • Molecules

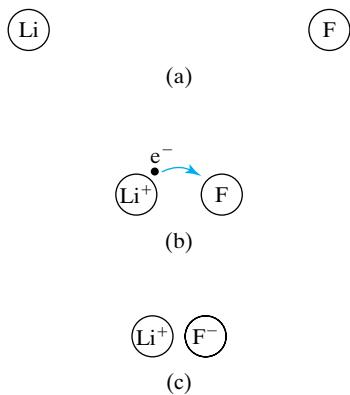


FIGURE 12.3

(a) When widely separated, the lithium and fluorine atoms are neutral. (b) When they approach one another, an electron can transfer from the lithium to the fluorine. (c) The resulting charged ions are strongly attracted and form a stable LiF molecule.

field and experiences no net force when placed in an applied field.* Equally important, the electron distribution in a noble atom cannot easily be distorted because the excited states are high above the ground state (about 20 eV in He). Thus, a noble atom, even when close to another atom, remains nearly spherical, and there is very little force between the two atoms.

This argument shows correctly why there is little force between a noble atom and any other atom, provided that they do not overlap. One could imagine that if the atoms were so close that they overlapped, there might be an attractive force; but even this is not the case: Once the atoms overlap, we need to take account of the quantized energy levels of the electrons, and we will see later that because the noble atom has all closed shells, there can be very little attraction, even when the atoms overlap.

Ionic Bonding

The simplest example of a pair of atoms that do form a strong bond is an alkali-halide pair, consisting of one alkali and one halogen atom. As we discussed in Section 10.7, the alkali metals (Li, Na, K, ...) have one loosely bound valence electron outside a closed shell. The halogens (F, Cl, Br, ...) have one vacancy in their outer shell and hence have large electron affinity. If, for example, the alkali Li approaches the halogen F, the F atom can capture the valence electron of the Li atom, as indicated schematically in Fig. 12.3. After the transfer of an electron from Li to F, the resulting ions are stable and spherical, since each now has a closed outer shell. The strong electrostatic attraction between these two oppositely charged stable spheres forms an **ionic bond**, which binds them together in a molecule of LiF.

Since both ions are spherically symmetric, they behave like point charges of opposite sign and form an *electric dipole*. The **dipole moment** p of such a pair of equal but opposite charges is defined to be

$$p = qd \quad (12.1)$$

where q is the magnitude of either charge and d is their separation. We have seen that a typical separation distance is $d \approx 0.2 \text{ nm}$, so an ionically bonded molecule should have a dipole moment of order

$$p = ed \sim (1.6 \times 10^{-19} \text{ C}) \times (2 \times 10^{-10} \text{ m}) = 3.2 \times 10^{-29} \text{ C} \cdot \text{m} \quad (12.2)$$

Measurements show that the dipole moments of alkali halide molecules are of this order, confirming that their bonding does indeed involve the transfer of an electron, of the kind illustrated in Fig. 12.3 (see Problem 12.3).

Covalent Bonding

When two identical atoms approach one another, there is no difference in electron affinities to cause the transfer of an electron as required for an ionic bond. Nonetheless, stable molecules such as H_2 , N_2 , and O_2 are known to exist. Furthermore, many other diatomic molecules, such as CO, have dipole moments that are much smaller than that predicted in (12.2) and thus cannot be ionically bonded. The bond that forms in all these cases involves a *sharing*,

*This second claim follows from Newton's third law: The spherical distribution exerts no force on any external charge; therefore, the external charge exerts no force on the distribution.

or *co-ownership*, of the valence electrons from both atoms, and is called a **covalent bond**.

We will discuss the covalent bond in detail in Section 12.4, but the main result of that discussion is easily summarized with the help of Fig. 12.4, which is a sketch of the probability density for the valence electrons around two atoms joined by a covalent bond. Part (a) shows the electron distributions around the two well separated atoms. In (b) the atoms are close together and a large proportion of the electron distribution is concentrated in a region between the two nuclei. The electrostatic attraction of the positive nuclei toward this concentration of negative charge is what bonds the atoms together.

Mixed Bonds

In a purely ionic bond an electron is completely transferred from one atom to another, and the dipole moment is given by (12.2). In a purely covalent bond the shared electrons are equally divided between the atoms, and the dipole moment is zero. While a few molecules (such as H_2 and O_2) are purely covalent, there are no molecules that are 100% ionic, and in most molecules the bonding is a combination of ionic and covalent. That is, there is a partial transfer of charge from one atom to the other and a partial sharing. This is signaled by a dipole moment that is not as large as expected for an ionic bond, but is still greater than zero.

Estimating Bond Strengths

In the next two sections we discuss in some detail the strengths of ionic and covalent bonds. Here we show that it is easy to get a rough estimate of their strengths. In particular, we can see why both kinds of bond lead to binding energies that are of order a few eV, as we saw in Table 12.1.

To estimate the binding energy of an ionic molecule, let us suppose that one electron is completely transferred from one atom to the other. The potential energy of the two resulting ions, at separation R_0 , is $U = -ke^2/R_0$, and if we substitute the observed value $R_0 \approx 0.2$ nm, we find that

$$U = -\frac{ke^2}{R_0} \approx -\frac{1.44 \text{ eV} \cdot \text{nm}}{0.2 \text{ nm}} \approx -7 \text{ eV} \quad (12.3)$$

We will see in the next section that this potential energy is the dominant contribution to the molecule's total energy. If we simply ignore all other contributions, we get the rough estimate $E \approx U \approx -7$ eV. That this is negative confirms that the molecule is a stable bound state, and the binding energy of the ionic molecule is (in this approximation)

$$B = -E \approx \frac{ke^2}{R_0} \approx 7 \text{ eV} \quad (12.4)$$

Comparing this value with the measured values in Table 12.1, we see that our estimate is a bit large. As we will see in Section 12.3, the terms that we have neglected in (12.4) reduce our value of B and give a result that agrees well with experiment. Nevertheless, the rough estimate (12.4) is definitely of the right order of magnitude.

One can get a similar estimate of the binding energy of a covalent molecule by making reasonable assumptions about the distribution of charge in this type of molecule (see Problem 12.11). Here, however, we will point out only

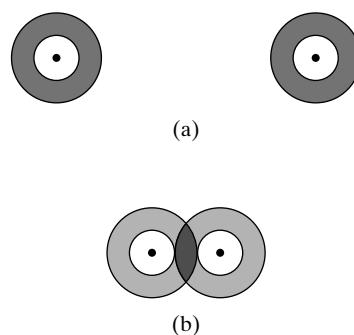
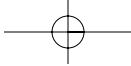


FIGURE 12.4

Schematic plot of the distribution of the outer electrons in two atoms that bond covalently. (a) The two separate atoms. (b) When the atoms form a covalent molecule, the wave functions for the outer electrons interfere constructively and produce a concentration of charge in the region between the two nuclei. The two dots show the positions of the two nuclei, and for clarity the distribution of inner electrons is omitted entirely.

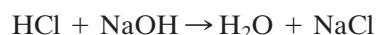


374 Chapter 12 • Molecules

that, as is clear from Table 12.1, the binding energy of covalent molecules is of the same order of magnitude as that of ionic molecules, and this similarity can be explained. The potential energy of the covalent molecule is the energy that results from concentrating one or two electrons between two positive charges, separated by a distance R_0 (as shown in Fig. 12.4). This potential energy is naturally of order (Problem 12.11) $U \approx -ke^2/R_0$, which is the same expression (12.3) that we obtained for ionic molecules,* and we conclude — just what is observed — that covalent and ionic molecules would have binding energies of comparable magnitudes.

Energy Released in Chemical Reactions

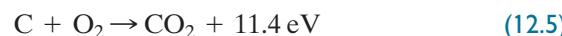
The binding energy of molecules manifests itself on a macroscopic scale in chemical reactions. A chemical reaction is simply a regrouping of the atoms in some initial set of molecules to form different molecules; for example,



Since the initial and final molecules generally have different binding energies, this regrouping results in the release or absorption of energy (usually in the form of heat or light). Binding energies are of order a few eV, so the energy released in an individual regrouping can itself be of this same order. On a macroscopic scale a few eV is a very small energy, but macroscopic amounts of matter contain enormous numbers of molecules, and the total energy released when a substantial amount of matter reacts can be very large, as the following example illustrates.

Example 12.2

Gasoline engines use the heat produced in the combustion of the carbon and hydrogen in gasoline. One of the important sources of energy in this combustion is the oxidation of carbon to form carbon dioxide:



where the 11.4 eV released comes from the increased binding energy of the CO_2 molecule as compared to that of the separate C and O_2 . Find the total energy released when 1 kg of carbon is oxidized. If this energy were used with 10% efficiency (a reasonable practical value) to drive a 1500-kg car up a hill, what elevation gain would it produce?

Because a mole of carbon is 12 grams, the number of carbon atoms in 1 kg is

$$N = \frac{1000 \text{ grams}}{12 \text{ grams/mole}} \times \left(6.0 \times 10^{23} \frac{\text{atoms}}{\text{mole}} \right) = 5.0 \times 10^{25} \text{ atoms}$$

*We should emphasize that this is only an order-of-magnitude argument. It may help to draw a parallel with the hydrogen atom, whose energy could be estimated as the potential energy of an electron and proton a distance a_B apart. This gives $E \approx -ke^2/a_B$, which is twice the correct answer but is certainly the right order of magnitude.

The oxidation of each carbon atom yields 11.4 eV, so the total energy released is

$$\begin{aligned} E &= N \times (11.4 \text{ eV}) = (5.0 \times 10^{25}) \times (11.4 \text{ eV}) \\ &= 5.7 \times 10^{26} \text{ eV} = 9.1 \times 10^7 \text{ J} \end{aligned} \quad (12.6)$$

This energy results from burning 1 kg of carbon in the reaction (12.5). Since this reaction is the main source of energy in the combustion of gasoline and carbon is the main constituent of gasoline, we can take (12.6) as a rough estimate of the energy released in burning 1 kg of gasoline. (For a more careful estimate, see Problem 12.9.)

If 10% of the energy (12.6) were used to raise a 1500-kg car through a vertical height h , then, since the gain in potential energy would be mgh , we can calculate h as

$$h = \frac{\text{gain in PE}}{mg} = \frac{9.1 \times 10^6 \text{ J}}{(1500 \text{ kg}) \times (9.8 \text{ m/s}^2)} \approx 620 \text{ m}$$

This is the elevation gain produced by 1 kg of gasoline. Since a car's gas tank holds at least 40 kg, we see that a tankful of gasoline can lift a car through a vertical gain of some 25,000 m (or 16 miles) — some three times the total height of Mount Everest! This startling conclusion emphasizes that gasoline is an extremely concentrated source of energy.*

12.3 The Ionic Bond

We have seen that an **ionic bond** is formed when an electron is transferred from one atom to another, producing a pair of oppositely charged ions that attract one another strongly. This occurs when the first atom has a low ionization energy and can lose an electron easily, while the second has high electron affinity and can bind an extra electron relatively well. We start this section by considering the specific case of NaCl and examine in detail the energy balance in the formation of an NaCl molecule.

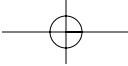
To begin, let us imagine that our Na and Cl atoms are far apart and electrically neutral (in which case they exert no force on one another). If we were to transfer an electron from the Na to the Cl while the atoms were still far apart, this transfer would actually *cost* us energy. The energy needed to remove an electron from Na is its ionization energy, $\text{IE(Na)} = 5.1 \text{ eV}$; the energy gained when Cl captures an electron is its electron affinity, $\text{EA(Cl)} = 3.6 \text{ eV}$. Thus the net cost of energy is

$$\Delta E = \text{IE(Na)} - \text{EA(Cl)} = 5.1 \text{ eV} - 3.6 \text{ eV} = 1.5 \text{ eV} \quad (12.7)$$

Clearly, the spontaneous transfer of an electron between the well-separated atoms is something that will not occur. However, this situation changes as the atoms approach one another, as we now show.

To understand what happens as the two atoms come closer together, consider the two graphs in Fig. 12.5. These show the energy, as a function of the

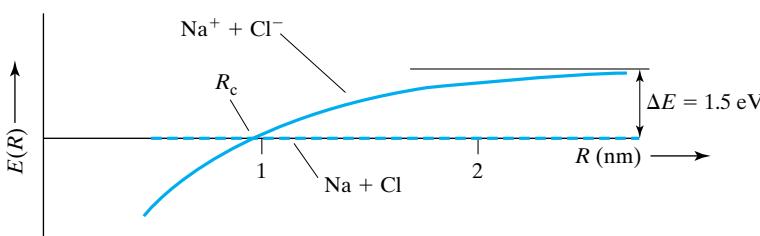
*We have assumed here that the car drives slowly up a steep incline, so that the work against rolling and air resistance is small compared to the work, mgh , against gravity. In the more usual context of a level road, the mgh term is zero and only rolling and air resistance remain; under these conditions a tankful of gasoline takes one several hundred miles (see Problem 12.5).



376 Chapter 12 • Molecules

FIGURE 12.5

Energy of sodium and chlorine as a function of their internuclear separation R . The dashed horizontal line is the energy of the two neutral atoms; the solid curve is that of the two ions Na^+ and Cl^- . When R is less than R_c , the two ions have lower energy than the two neutral atoms.



distance R between the two nuclei, both for the neutral pair $\text{Na} + \text{Cl}$ (dashed), and for the ions $\text{Na}^+ + \text{Cl}^-$ (solid curve). We have chosen the zero of energy as the energy of the two neutral atoms when far apart (and at rest). Since the neutral atoms exert no force on one another, their energy does not change as we move R inward, and the corresponding graph (shown as the dashed line) is a constant:

$$E(R) = \text{constant} = 0 \quad [\text{neutral atoms}]$$

Suppose instead that while the atoms are far apart ($R = \infty$) we transfer an electron from one to the other, and then slowly move them together. While they are still far apart, the energy of the ions is higher than that of the neutral atoms by the amount $\Delta E = 1.5 \text{ eV}$ calculated in (12.7). However, the ions are oppositely charged, and as they move closer, their potential energy, $-ke^2/R$, comes into play, so their total energy is

$$E(R) = \Delta E - \frac{ke^2}{R} \quad [\text{Na}^+ + \text{Cl}^-] \quad (12.8)$$

This energy is shown as the solid curve in Fig. 12.5 and decreases steadily as the ions get closer.

At the critical separation shown as R_c in Fig. 12.5, the energy of the two ions Na^+ and Cl^- has dropped to that of the neutral pair, $\text{Na} + \text{Cl}$. This separation is easily calculated by setting $E(R)$ in (12.8) equal to zero, to give

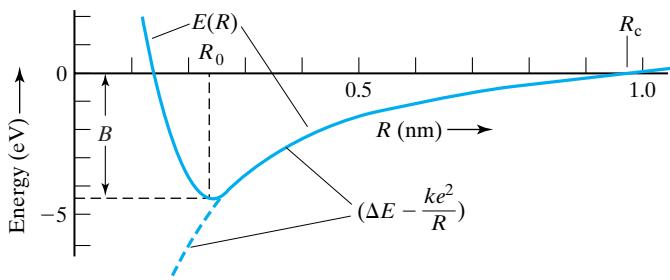
$$R_c = \frac{ke^2}{\Delta E} \quad (12.9)$$

For the case of NaCl , with $\Delta E = 1.5 \text{ eV}$, this gives

$$R_c = \frac{ke^2}{\Delta E} = \frac{1.44 \text{ eV} \cdot \text{nm}}{1.5 \text{ eV}} = 0.96 \text{ nm}$$

Once R is less than R_c , the ions have less energy than the neutral atoms; that is, the potential energy $-ke^2/R$ of the ions more than offsets the cost ΔE of transferring the electron, and the transfer is energetically favored. To predict the probability that the transfer actually occurs would require a detailed quantum calculation, but for our purposes, it is sufficient to note that the common occurrence of ionic molecules shows that the transfer does have significant probability.

Once the electron is transferred, the two ions are strongly attracted and can bind together to form a molecule. Their energy does not continue to decrease indefinitely as R approaches zero. Rather, once the two ions overlap



appreciably, the force between them becomes repulsive, and the energy $E(R)$ rises as shown in Fig. 12.6. One obvious source of this repulsion is the increasing force between the two positive nuclei as they get closer. Another important contribution is due to the Pauli exclusion principle: Electrons in the same region cannot occupy the same quantum state, and once the atoms overlap, some of the electrons must move to higher levels, increasing $E(R)$ further.

The energy $E(R)$ of the two ions has a minimum at the separation shown as R_0 in Fig. 12.6, and the two ions are in stable equilibrium at this separation. Therefore, R_0 is the center-to-center distance of the two atoms in the lowest state of the stable molecule, and the corresponding value of the energy, $E(R_0)$, is the ground-state energy of the molecule.* That is, $E(R_0) = -B$, where B is the binding energy of the molecule. The observed values of these two parameters for NaCl are

$$R_0 = 0.24 \text{ nm} \quad \text{and} \quad B = 4.2 \text{ eV}$$

Without a detailed calculation of the repulsive contributions to $E(R)$, we cannot actually *predict* these values of R_0 and B . However, one can see in Fig. 12.6 that when R is close to R_0 , the repulsive contribution to $E(R)$ is still quite small; that is, near $R = R_0$ the exact $E(R)$, shown by the solid curve, is close to the value $(\Delta E - ke^2/R)$ shown by the dashed curve. This suggests that we would get good approximations for $E(R_0)$ and B by ignoring the difference between these two curves; that is, we can approximate

$$B = -E(R_0) \approx \frac{ke^2}{R_0} - \Delta E \quad (12.10)$$

This estimate for B is easily understood: The term ke^2/R_0 is the electrostatic binding energy of the two ions and is reduced by the energy ΔE that was needed to transfer the electron and form the ions.

In Section 12.2 we made the rough estimate

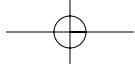
$$B \approx \frac{ke^2}{R_0} \quad (12.11)$$

Comparing this with the new estimate (12.10), we see that this earlier approximation simply ignored the transfer energy ΔE , and was therefore an overestimate, as we noted in Section 12.2. In fact, the improved approximation (12.10)

FIGURE 12.6

The general behavior of the energy $E(R)$ of the two ions Na^+ and Cl^- is shown by the solid curve. As long as the two ions do not overlap ($R \gtrsim 0.3 \text{ nm}$) the energy is just $E(R) = \Delta E - ke^2/R$. Once the ions overlap, repulsive forces quickly dominate and $E(R)$ climbs steeply up again. The minimum of $E(R)$ occurs at R_0 , and $E(R_0) = -B$, where B is the molecule's binding energy.

*We are still ignoring any kinetic energy of the two atoms. As we discuss in Section 12.6, there is a small, zero-point kinetic energy, and the ground state is actually slightly above $E(R_0)$.



378 Chapter 12 • Molecules

is still a slight overestimate since it ignores the repulsive contribution to $E(R)$ and this contribution necessarily reduces B .

Example 12.3

Given the known values of R_0 and ΔE for NaCl, estimate the dissociation energy B of the NaCl molecule, using both of the approximations (12.11) and (12.10).

Since $R_0 = 0.24 \text{ nm}$, the rough estimate (12.11) gives

$$B \approx \frac{ke^2}{R_0} = \frac{1.44 \text{ eV} \cdot \text{nm}}{0.24 \text{ nm}} = 6.0 \text{ eV}$$

while the improved estimate (12.10) gives

$$B \approx \frac{ke^2}{R_0} - \Delta E = (6.0 - 1.5) \text{ eV} = 4.5 \text{ eV}$$

As expected, both approximations overestimate the observed value of 4.2 eV. The first is within 40%, and the second a respectable 7%.

Valence

The NaCl molecule is formed by the transfer of one electron from an Na atom to a Cl atom, to give the closed-shell ions Na^+ and Cl^- . Similar ionic molecules, involving transfer of one electron, can be formed by combining any alkali atom (closed-shell-plus-one, or group I) with any halogen (closed-shell-minus-one, or group VII).

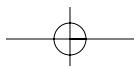
Ionic molecules that involve the transfer of more than one electron are also possible. For example, an atom like Mg from group II, has two electrons outside closed shells and can lose these two electrons to become the closed-shell ion Mg^{2+} ; meanwhile an atom like O from group VI is two electrons short of a closed shell, so can bind two extra electrons to form the closed-shell O^{2-} . Thus, Mg and O can combine to form the ionic molecule MgO by transfer of two electrons. Similarly, an atom such as Al from group III can lose three electrons, and an atom such as N from group V can gain three; thus, one can form the ionic molecule AlN by transfer of three electrons from Al to N.

There are also many possible ionic molecules involving more than two atoms. For example, one can form the ionic molecule MgF_2 , in which the Mg atom loses two electrons (to become Mg^{2+}) while each of the two F atoms gains one electron (to become F^-). In Table 12.2 we show the nine possibilities

TABLE 12.2

The nine possible molecules formed by combining the positive ions Na^+ , Mg^{2+} , or Al^{3+} with the negative ions F^- , O^{2-} , or N^{3-} . Numbers in parentheses show the valence of each element concerned.

F⁻ (1)	O²⁻ (2)	N³⁻ (3)
$\text{Na}^+(1)$	NaF	Na_2O
$\text{Mg}^{2+}(2)$	MgF_2	MgO
$\text{Al}^{3+}(3)$	AlF_3	Al_2O_3



for the ionic molecules formed from Na, Mg, or Al, each in combination with one of the elements F, O, or N.

When an atom can form ionic molecules, we define its **valence** as the number of electrons it gains or loses in forming such a molecule. Elements from groups I and VII (closed-shell \pm one) have valence 1; elements from groups II and VI (closed-shell \pm two) have valence 2; and so on. Knowing the valences of two elements that form an ionic molecule, one can immediately predict the proportions in which the atoms combine. For example, Mg has valence 2 (closed-shell ion Mg^{2+}) and N has valence 3 (closed-shell ion N^{3-}). To retain overall neutrality, we must combine three Mg^{2+} ions with every two N^{3-} ions, and the resulting compound is Mg_3N_2 , as shown in Table 12.2. In general, the proportion of constituent atoms in any ionic molecule is just the reciprocal of the ratio of their valences.

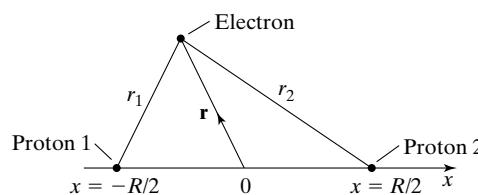
12.4 The Covalent Bond

Most diatomic molecules that are not bound ionically are bound instead by the **covalent bond**. As already described, the covalent bond involves a concentration of one or more electrons from each atom in the region between the two nuclei. To understand how this concentration occurs, we begin by considering the simplest of all molecules, the H_2^+ molecular ion, which consists of two 2 protons and a single electron.* Just as an understanding of the one-electron atom (hydrogen) helped us to build a theory of multielectron atoms in Chapter 10, so we will see here that an understanding of the one-electron molecule (H_2^+) will help us to understand the covalent bonding of multielectron molecules such as H_2 , O_2 , and H_2O .

The H_2^+ Molecular Ion

In H_2^+ the single electron moves in the field of the two protons. Because the protons are so heavy, they move very little compared to the electron, and it is a good approximation to treat them as stationary. (This is the same approximation used in Chapters 8 and 10 to discuss atoms.) Within this approximation, we must solve the Schrödinger equation for the single electron in the field of the two fixed protons and find its lowest allowed energy E . Just as with ionic molecules, this energy will depend on the internuclear distance R ; that is, $E = E(R)$. If $E(R)$ has a minimum at some separation R_0 , a stable molecule exists, with bond length R_0 and energy $E(R_0)$.

Before trying to solve the Schrödinger equation, we must choose a system of coordinates. A convenient choice is to put the two protons on the x axis at positions $x = \pm R/2$, as shown in Fig. 12.7. The electron's position is denoted



Linus Pauling
(1901–1994, American)

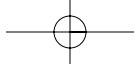


One of the founding fathers of quantum chemistry, Pauling applied the principles of quantum mechanics to the bonding of atoms in molecules, and his book, *The Nature of the Chemical Bond* (1939), was one of the most influential texts of the period. He was one of the first to suggest that protein molecules are arranged in a spiral. Pauling won the Nobel Prize in chemistry in 1954 and the Nobel Peace Prize, for his fight against nuclear weapons, in 1962.

FIGURE 12.7

The H_2^+ molecule consists of two protons and a single electron. We treat the protons as fixed at $x = \pm R/2$ on the x axis. The electron's position relative to the origin is denoted by \mathbf{r} . Its electrostatic potential energy is determined by the distances to the two protons, shown as r_1 and r_2 .

*The H_2^+ ion is readily formed in hydrogen gas by an electric discharge, which can strip an electron from a neutral H_2 molecule. In isolation, H_2^+ is perfectly stable. In practice, it eventually picks up a free electron and reverts to neutral H_2 . Nevertheless, it survives quite long enough for an accurate determination of its properties.



380 Chapter 12 • Molecules

by \mathbf{r} , and we must find the wave function $\psi(\mathbf{r})$ that gives the minimum total energy for the molecule with a given separation R .

A direct analytic solution of the Schrödinger equation for the electron in H_2^+ is possible, but it is rather complicated and not especially illuminating. For our purposes, it is more convenient to seek an approximate solution; and to this end, we begin by supposing that the two protons are far apart — specifically, that R is much larger than the Bohr radius a_B . Under these conditions the solutions of the Schrödinger equation are easy to see. If the electron is close to proton 1, the effect of the distant proton 2 is small and the lowest possible state is just the ground state for an electron bound to proton 1. The corresponding wave function, which we denote by $\psi_1(\mathbf{r})$, is the familiar $1s$ wave function for a hydrogen atom centered on proton 1:

$$\psi_1(\mathbf{r}) = Ae^{-r_1/a_B} \quad (12.12)$$

where r_1 is the distance from proton 1 to the electron. There is a second state with exactly the same energy, in which the electron is bound to proton 2, with wave function

$$\psi_2(\mathbf{r}) = Ae^{-r_2/a_B} \quad (12.13)$$

We have sketched these two wave functions in Fig. 12.8, which shows the values of $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ along the x axis (that is, the line joining the two protons).

The two wave functions ψ_1 and ψ_2 satisfy the Schrödinger equation with the same energy; that is, they are degenerate. Therefore, any linear combination,

$$\psi = B\psi_1 + C\psi_2 \quad (12.14)$$

(for any two constants B and C) also solves the Schrödinger equation with this same energy. This wave function is a superposition of a state where the electron is bound to proton 1 and another where it is bound to proton 2, and — in principle at least — any such state is a possible state of our system. Two particular such states that we will find to be especially important are*

$$\psi_+ = \psi_1 + \psi_2 \quad (12.15)$$

and

$$\psi_- = \psi_1 - \psi_2 \quad (12.16)$$

These two wave functions are sketched in Fig. 12.9(a). Their characteristic property is that they represent states where the electron is distributed equally around both protons. To see this, note that, as is clear from Fig. 12.9(a),†

$$\psi_+(x) = \psi_+(-x) \quad \text{and} \quad \psi_-(x) = -\psi_-(-x) \quad (12.17)$$

*If we want these functions to be normalized, we should include an overall normalization factor; since this factor does not affect our discussion here, we omit it (but see Problems 12.31 and 12.56).

†For simplicity, we focus here on the wave functions on the x axis [and for brevity write them as $\psi_{\pm}(x)$]. The corresponding symmetry for arbitrary points $\mathbf{r} = (x, y, z)$ is this: $\psi_+(\mathbf{r}) = \psi_+(-\mathbf{r})$ and $\psi_-(\mathbf{r}) = -\psi_-(-\mathbf{r})$.

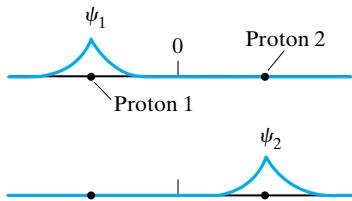


FIGURE 12.8

Wave functions for the one electron in the H_2^+ molecule with the two protons far apart ($R \gg a_B$). The plots show the values of ψ on the internuclear axis. The function ψ_1 describes a state in which the electron is bound in the $1s$ state around proton 1 and is therefore affected very little by the distant proton 2; ψ_2 is the corresponding state with the electron bound to proton 2.

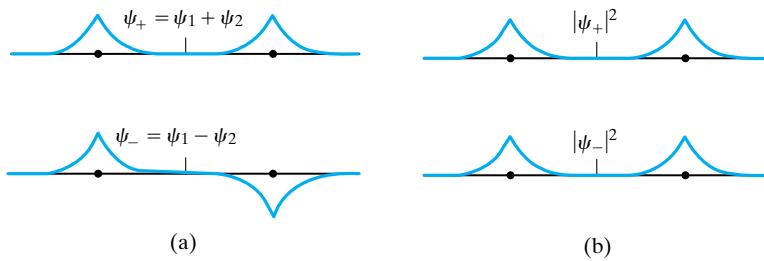


FIGURE 12.9

(a) The wave functions ψ_+ and ψ_- for the electron in H_2^+ when the two protons are far apart. The plots show values of $|\psi_{\pm}|$ along the internuclear axis. **(b)** Corresponding plots of the electron's probability density $|\psi_{\pm}|^2$ (which are identical as long as the protons are far apart).

that is, $\psi_+(x)$ is an even function of x and $\psi_-(x)$ is an odd function. Therefore, for either wave function, the probability density $|\psi|^2$ is the same at any point x as at the point $-x$:

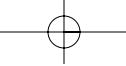
$$|\psi(x)|^2 = |\psi(-x)|^2 \quad (12.18)$$

with ψ equal to either ψ_+ or ψ_- . This is illustrated in Fig. 12.9(b) and means that the electron distribution at any point near proton 1 is the same as at the corresponding point near proton 2.

The individual wave functions ψ_1 and ψ_2 are often called **atomic orbitals** since they describe states where the electron forms an atom with one of the nuclei and is unaffected by the other. [Recall that “orbital” is simply an alternative name for the spatial wave function $\psi(\mathbf{r})$ for any one electron; that is, an orbital specifies the orbital motion — but not the spin state — of one electron. In an atom an orbital is specified by the quantum numbers n, l, m .] The combinations ψ_+ and ψ_- are called **molecular orbitals** since they describe states where the electron is associated with *both* nuclei and belongs to the molecule as a whole. We will see that it is the molecular orbitals ψ_+ and ψ_- that become the stationary states of H_2^+ when the two protons come close enough to form a stable molecule.

When the two protons approach one another, the wave function $\psi_1(\mathbf{r})$ overlaps proton 2 and vice versa, and the electron's motion is influenced by both protons. The simple atomic wave functions ψ_1 and ψ_2 of Eqs. (12.12) and (12.13) are no longer solutions of the complete Schrödinger equation (which includes the potential energy due to both protons). Nevertheless, we can show that the combinations $\psi_{\pm} = \psi_1 \pm \psi_2$ remain reasonable approximations. Any correct solution should reflect the symmetry of the system, which has two identical force centers at $x = \pm R/2$; that is, the wave functions for the stationary states of H_2^+ should have the symmetry property (12.18). Now, as we have just seen, the wave functions $\psi_{\pm} = \psi_1 \pm \psi_2$ do have this symmetry. Furthermore, these functions, being combinations of ψ_1 and ψ_2 reflect the presence of both protons. This suggests that the functions ψ_+ and ψ_- could be reasonable approximations to correct solutions of the Schrödinger equation, even though ψ_1 and ψ_2 separately are not. Using a more advanced argument (called the variational method), one can show that this suggestion is correct. Here we will simply accept that ψ_+ and ψ_- are reasonable approximate solutions, and use them to deduce the general behavior of the energy of the H_2^+ molecule for the two states concerned.

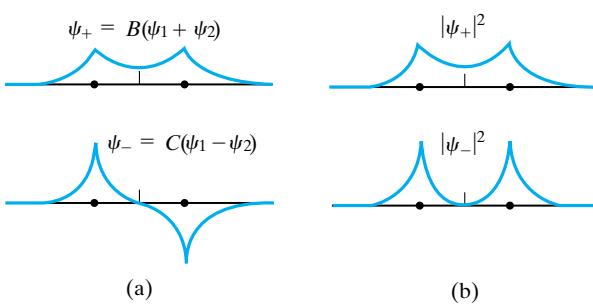
Once the two protons are within one or two Bohr radii, the behavior of ψ_{\pm} is as sketched in Fig. 12.10(a). The corresponding probability densities $|\psi_+|^2$ and $|\psi_-|^2$ are shown in Fig. 12.10(b), and the important difference between the two states is immediately clear: The probability density $|\psi_+|^2$ is nonzero at, and near to, the origin; in fact, it is larger in this region than either $|\psi_1|^2$ or $|\psi_2|^2$ separately (see Problems 12.31 and 12.56). This means that for the state ψ_+ , the electron distribution is enhanced in the region between the two protons —



382 Chapter 12 • Molecules

FIGURE 12.10

(a) Sketch of the wave functions ψ_+ and ψ_- for the electron in the H_2^+ molecule when the distance R between the two protons is comparable to the size of an H atom. At the origin, ψ_+ is larger than either ψ_1 or ψ_2 whereas ψ_- is exactly zero. (The factors B and C are normalization constants; C is a little larger than B — see Problem 12.31 — and this is why the peaks of ψ_- are a little taller than those of ψ_+) (b) The corresponding probability densities. (As in Figs. 12.8 and 12.9, all plots show values along the internuclear axis.)



(a) (b)

precisely where it is strongly attracted by both protons. This lowers the molecule's energy (as compared to the energy when R is large) and causes the H_2^+ molecule to bind. By contrast, the wave function ψ_- is exactly zero at the origin. Therefore, $|\psi_-|^2$ is small in the region between the two protons, and the energy is not low enough to bind in this state.

In Fig. 12.11 we show schematic two-dimensional pictures of the states $\psi_{\pm} = \psi_1 \pm \psi_2$. The left side of each picture represents the state while the atoms are well separated and the distribution consists of two distinct spherical clouds.* The right side of each picture represents the distributions when the wave functions ψ_1 and ψ_2 begin to overlap. In the state ψ_+ the two functions ψ_1 and ψ_2 interfere constructively in the region of overlap and concentrate the probability density there. In the state ψ_- , they interfere destructively and reduce the corresponding density.

In Fig. 12.12 we have plotted the energy $E(R)$ of the H_2^+ molecule, as a function of the distance R between the two protons, for both of the states ψ_+ and ψ_- . The zero of energy has been chosen as the energy when the electron is bound to one of the protons and the other proton is far away ($R \rightarrow \infty$). As we have just argued, the energy $E_+(R)$ of the **bonding orbital** ψ_+ decreases as the protons get closer, until it reaches a minimum at a separation R_0 . Once R is less than R_0 , the repulsion of the two protons is dominant and $E_+(R)$ increases rapidly as R approaches zero. The important point is that, just as with the ionic molecules discussed in Section 12.3, the energy has a minimum at a separation R_0 , and a stable molecule can form, with $R = R_0$ and binding energy $B = -E_+(R_0)$. The calculated values,

$$R_0 = 0.11 \text{ nm} \quad \text{and} \quad B = 2.7 \text{ eV}$$

agree perfectly with experiment.

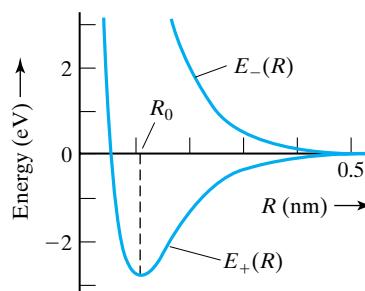


FIGURE 12.11

Schematic representation of the behavior of the molecular orbitals $\psi_{\pm} = \psi_1 \pm \psi_2$ of H_2^+ as the nuclei approach one another. The plus and minus signs show the signs of the wave functions. When the nuclei are far apart, both distributions consist of two separate spherical clouds. Once the clouds overlap, the wave functions ψ_1 and ψ_2 interfere constructively in the state ψ_+ and produce an enhanced density between the two nuclei; in the state ψ_- they interfere destructively and give a reduced density.

The energy of the H_2^+ molecule as a function of the distance R between the two protons. The curve $E_+(R)$ is the energy of the “bonding state” ψ_+ and $E_-(R)$ is that of the antibonding state ψ_- .

*We should emphasize that these figures are highly schematic. The distribution of ψ_1 or ψ_2 is not a clearly defined sphere but, rather, drops continuously from a maximum at the nucleus and reaches zero only at infinity. Thus the region of overlap is not really as sharply defined as our cartoons suggest.

The curve $E_-(R)$ in Fig. 12.12 increases steadily as R decreases and has no minimum; so the H_2^+ molecule has no stable bound state corresponding to the **antibonding orbital** ψ_- .

The Neutral H_2 Molecule

Armed with a theory of the one-electron molecule, H_2^+ , we can now understand many simple multielectron molecules. We consider first the neutral H_2 molecule with its two electrons. Just as in our discussion of multielectron atoms in Chapter 10, we start by ignoring the repulsion between the electrons. In this approximation each electron moves independently in the field of two protons, and its lowest level is just that of the bonding orbital ψ_+ of the H_2^+ molecule, with energy -2.7 eV . The ground state of H_2 , with its two electrons, is obtained by putting both electrons into this lowest level, to give a total energy of -5.4 eV , and hence binding energy $B = 5.4 \text{ eV}$. This value is an overestimate since it ignores the electrostatic repulsion of the two electrons. Nevertheless, it is reasonably close to the observed value $B = 4.5 \text{ eV}$.

We found the ground state of H_2 by giving both electrons the same wave function ψ_+ . According to the Pauli exclusion principle, this is possible only if the two electrons occupy different spin states; that is, the ground state of the H_2 molecule must have the two electrons' spins antiparallel and hence have $s_{\text{tot}} = 0$ — a prediction that is fully confirmed by experiment. If we try to assemble an H_2 molecule from two H atoms whose electron spins are *parallel*, then one electron can go into the bonding orbital ψ_+ , but the exclusion principle requires that the other go into the antibonding orbital ψ_- or some other higher level. With the second electron in ψ_- , it is found that the total energy $E(R)$ has no minimum, and there is no corresponding molecule; if the second electron goes to another, higher level, a molecule *may* form, but with much higher energy — that is, it forms an excited state (as we discuss in Section 12.6).

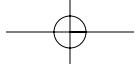
The ground state of the H_2 molecule typifies the main features of all covalent bonds. Each atom contributes one electron to the bond. The two electrons have their spins antiparallel, which allows them both to occupy the same bonding orbital with a large density between the two nuclei. The resulting “electron pair” produces a strong bond between the two atoms, but we will see that because of the exclusion principle, they cannot be joined by a third electron. It is the strong bond produced by a shared pair of electrons with antiparallel spins and their inability to accept a third electron that are the essential characteristics of the covalent bond.

Saturation of the Covalent Bond

One of the most striking features of many molecular bonds is the phenomenon of **saturation**: Molecules containing given elements usually contain some definite number of atoms and cannot bind any additional atoms. For example, hydrogen forms the stable molecule H_2 , as we have just seen, but a stable H_3 molecule does not exist.

To understand why this is true, we have only to recall part of our discussion of multielectron atoms in Chapter 10. In the two-electron atom, He, both electrons can occupy the lowest ($1s$) orbital and are very tightly bound. In the three-electron atom, Li, two of the electrons can occupy the lowest orbital, but the exclusion principle requires the third electron to occupy the next ($2s$) orbital, which is much less tightly bound.

Almost exactly the same argument applies to the molecules H_2 and H_3 : In H_2 both electrons can occupy the lowest molecular orbital (the bonding



384 Chapter 12 • Molecules

orbital ψ_+), and this molecule is tightly bound. In H_3 there is again an orbital of lowest energy, with the probability density concentrated inside the triangle formed by the three protons. However, as with the Li atom, only two of the three electrons can occupy this lowest orbital, and the third must occupy a higher orbital.

This qualitative argument shows clearly that an H_3 molecule, if it is bound at all, should be less tightly bound than H_2 . To decide whether or not the H_3 molecule (or indeed the Li atom) actually is bound requires a detailed calculation of the energy levels concerned. In the case of the Li atom, we know that it *is* bound, although with much lower ionization energy than He. In the case of the H_3 molecule, it turns out that the higher levels are so much higher that H_3 is not bound. The big difference between the Li atom and the H_3 molecule results from several effects, one of the most important differences concerns the three protons in either system: In the Li atom, the three protons are tightly bound inside the nucleus by the strong nuclear force; as far as the atom is concerned, the proton–proton repulsion is effectively switched off. In the hypothetical H_3 molecule, the proton–proton repulsion raises the energy by so much (several tens of eV) that the molecule is unstable.

If we consider possible larger hydrogen molecules, such as H_4 , this situation only becomes worse. Therefore, there are no hydrogen molecules beyond the familiar, diatomic H_2 , and the bonding between n hydrogen atoms “saturates” at $n = 2$.

Covalent Bonding of Multielectron Atoms

The formation of molecules from multielectron atoms is not as complicated as one might at first fear. This is because it is only the outer, or valence, electrons of each atom that overlap enough to have an important role in the molecular bond. A closed-shell-plus-one atom such as Li (with configuration $1s^22s^1$) or Na ($1s^22s^22p^63s^1$) can, for our purposes, be regarded as a single electron in an s state outside an inert positive “core.” This suggests that these atoms should behave very much like hydrogen, the main difference being that the one valence electron is in a state concentrated at much larger radius (being $n = 2, 3, \dots$). Thus, we would anticipate that a lithium atom could bond covalently with hydrogen to form LiH, or with a second lithium to form Li_2 . The molecule Li_2 should be bigger than LiH, which should be bigger than H_2 , and because of the greater separations, the binding energies should be correspondingly smaller. All of these predictions are confirmed by the experimental data shown in Table 12.3.

Atoms with more than one valence electron are naturally more complicated, but here again the situation is not as bad as one might fear. Consider, for example, the fluorine atom, with its seven valence electrons and configuration



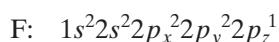
TABLE 12.3

Bond lengths (R_0) and binding energies (B) of the covalent molecules H_2 , LiH, and Li_2 .

Molecule	R_0 (nm)	B (eV)
H_2	0.074	4.5
LiH	0.16	2.4
Li_2	0.27	1.0

Section 12.4 • The Covalent Bond **385**

Recall that there are three independent $2p$ wave functions, or orbitals, which we can take to be the functions introduced in Section 8.8, labeled $2p_x$, $2p_y$, and $2p_z$, and we can rewrite the electron configuration (12.19) in more detail as

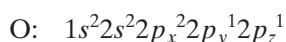


(It does not matter which of the three $2p$ orbitals gets the single electron; to be definite, we suppose that it is the $2p_z$.) The important point is that the orbitals $2s$, $2p_x$ and $2p_y$ are occupied by two electrons each. The electrons in each orbital must therefore have their spins antiparallel; that is, they form an electron pair that cannot participate in a covalent bond. Thus, the fluorine atom has just one electron available to share in a covalent bond with another atom. In other words, fluorine is, in this respect, similar to hydrogen. We would expect it to bond covalently with H to form HF, or with another F atom to form F_2 — and both of these molecules are indeed observed.

The oxygen atom has one less electron than fluorine, with the configuration



Two of the electrons in the $2p$ level must occupy the same orbital ($2p_x$, say) but as we argued in Section 10.7, the other two will go, one each, into the two remaining orbitals,* $2p_y$ and $2p_z$. That is, the detailed configuration of the oxygen atom is



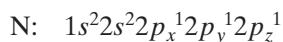
This leaves *two* unpaired electrons that are free to be shared in covalent bonds with other atoms. Thus, oxygen can form a molecule like H_2O , in which the O atom forms one covalent bond with each of the two H atoms. It can also form two covalent bonds, with a second O atom to make the doubly bonded molecule O_2 . Because O_2 is doubly bonded, we would expect it to be more strongly bound than the singly bonded F_2 , and this prediction is confirmed by the observed binding energies given in Table 12.4.

TABLE 12.4

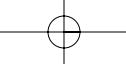
Binding, or dissociation, energies of the singly, doubly, and triply bonded molecules F_2 , O_2 , and N_2 .

Molecule	B (eV)
F_2	1.6
O_2	5.1
N_2	9.8

The nitrogen atom has one less electron than oxygen and has the configuration



*Recall that with this arrangement, the two electrons are farther apart (than if both occupied the same orbital) and have slightly lower energy because their Coulomb repulsion is reduced.



386 Chapter 12 • Molecules

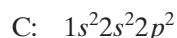
Since it has *three* unpaired electrons that are available for sharing with other atoms, it can, for example, form one bond with each of three H atoms to make an ammonia molecule NH_3 . It can also form three bonds with a second N atom to make the triply bonded N_2 . Since N_2 has three bonds, it is very well bound, as seen in Table 12.4; and for this reason, N_2 gas is chemically very inactive.

Valence

In the context of covalent bonding, we define the **valence** of an atom as the number of electrons that it can share with other atoms in covalent bonds; H, Li, and F all have valence 1, while O has valence 2, and N has valence 3. In the case of ionic molecules such as LiF, we defined the valence of an atom as the number of electrons the atom can gain or lose to form the ionic molecule. (F has valence 1 because it *gains* one electron to become F^- , whereas Li has valence 1 because it *loses* one electron to become Li^+ .) Since many atoms can bond both covalently and ionically, it is important to note that in all the simple examples discussed here, our two definitions give the same value for the valence. For instance, a closed-shell-plus-one atom such as Li or Na can *lose* its outer electron to form an ionic bond, and it can *share* the outer electron to form a covalent bond. Either way, its valence is 1. Similarly, a closed-shell-minus-one atom such as F or Cl can *gain* an electron (which fills the one vacancy in its outer shell) to form an ionic bond; but as we have just seen, it can *share* the one unpaired electron in its outer shell to form a covalent bond. Either way, its valence is also 1.

On the basis of what has been said here, it should be clear that a molecule like LiF could bond ionically *or* covalently and that (at the level of our qualitative treatment) we have no way to predict which mechanism will prevail. By measuring the dipole moment, as discussed in Section 12.2, one can find the extent to which a given molecule is ionically or covalently bonded. In the case of the alkali halide molecules (such as LiF) it is found that the bond is predominantly ionic, but in many other molecules the bond is found to be a combination of both types.

We should emphasize that we have focused our discussion on the simplest molecules and that many molecules are much more complicated. Perhaps the most important example of the complications that can occur, even with fairly simple atoms, concerns carbon. The ground state of carbon has the configuration

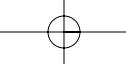


which suggests that carbon could contribute two electrons to covalent bonds and should have valence 2, as it does in CO. However, carbon usually has valence 4, as is testified by the existence of such molecules as CH_4 (methane) and CO_2 . The explanation of this apparent contradiction is actually quite simple: The 2s and 2p levels are close together, and only a small energy is needed to raise one electron to give the configuration



In this configuration carbon has four electrons available for sharing in covalent bonds, and the energy gained by forming four bonds amply repays the small energy needed to achieve the configuration (12.20).

Carbon can use its four valence electrons to form bonds in many different ways. For example, it can form the molecule CH_4 , as shown schematically

Section 12.5 • Directional Properties of Covalent Bonds **387**

in Fig. 12.13(a). However, it can also form the molecule C_2H_4 (ethylene), in which each carbon is joined to two hydrogens but is also doubly bonded to the other carbon, as shown in Fig. 12.13(b). Another example of the numerous *hydrocarbon* molecules is C_6H_6 (benzene), in which the six carbons are joined in a ring of single and double bonds; this leaves room for one hydrogen to attach to each carbon, as shown in Fig. 12.13(c). Because carbon can bond in so many ways to other carbon atoms and to other elements, such as H, N, O, and many more, there is a vast array of different carbon compounds. Since many of them play an important role in organisms, all carbon-containing molecules (with a few exceptions such as CO and CO_2) have come to be called **organic molecules**.

12.5 Directional Properties of Covalent Bonds*

*This section can be omitted without serious loss of continuity.

In this section we describe how the properties of atomic wave functions determine the shapes of molecules. To begin, we review briefly the formation of a simple diatomic molecule like H_2 .

As we saw in Section 12.4, the two electrons in H_2 both occupy the bonding orbital, which corresponds to the combination $\psi_+ = \psi_1 + \psi_2$ of the $1s$ wave functions for each separate atom,

$$\psi_1 = Ae^{-r_1/a_B} \quad \text{and} \quad \psi_2 = Ae^{-r_2/a_B}$$

Since these functions are both positive, they interfere constructively in their region of overlap (for the orbital ψ_+), as was indicated schematically in Fig. 12.11(a). This enhances the charge density between the two nuclei and lowers the energy of the molecule.

Similar considerations apply to any diatomic molecule in which the valence electrons are all in s states, and Fig. 12.11(a) could equally have represented such molecules as Li_2 and LiH . If we consider instead such atoms as N, O, or F, whose valence electrons are in p states, the situation is more interesting. Recall that p states have three possible wave functions, which we can take to be the functions p_x , p_y , and p_z introduced in Section 8.8 and illustrated in Fig. 8.21 (which we have redrawn here as Fig. 12.14). The important point about each of these p states is that its probability density is concentrated in two lobes along one of the three coordinate axes. For example, the $2p_z$ wave function of hydrogen is Eq. (8.94)

$$\psi_{2p_z} = Aze^{-r/2a_B} \quad (12.21)$$

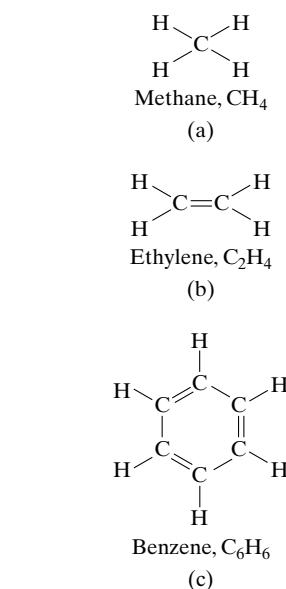
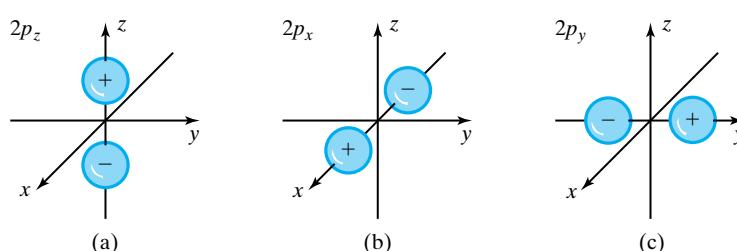
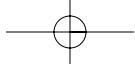


FIGURE 12.13

Schematic views of three hydrocarbon molecules: CH_4 (methane), C_2H_4 (ethylene), and C_6H_6 (benzene). Each short line represents a covalent bond formed by a shared pair of electrons. The actual molecules are, of course, three-dimensional; for example, the four H atoms in CH_4 are arranged symmetrically at the four corners of a tetrahedron with the carbon at its center. In benzene the six bonds between the carbon pairs are perhaps better thought of as all being equal mixtures of single and double bonds.

FIGURE 12.14

Each of the three wave functions p_x , p_y , p_z has two lobes where the probability density is largest. (These pictures show the contours where the density is 75% of its maximum value, for the $2p$ wave functions of hydrogen.) The plus and minus signs indicate that the wave function is positive in one lobe and negative in the other.



388 Chapter 12 • Molecules

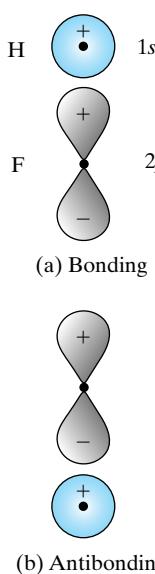


FIGURE 12.15

Wave functions for the HF molecule. (a) We obtain the bonding orbital by adding the wave functions for the 1s electron in H and the 2p_z electron in F, provided that the H atom approaches the positive lobe of the 2p_z function. (The 2p_z wave function in F is shown here as a figure eight to emphasize that both lobes belong to a single atomic wave function.) (b) If the H atom approaches from the opposite direction, adding the two wave functions gives an antibonding orbital. For clarity, the two atoms are shown with their wave functions not yet overlapping.

This gives maximum probability on the z axis and zero probability in the xy plane (where z = 0). For our present purposes, it is important to note that the function (12.21) is positive in its upper lobe (where z > 0) and negative in the lower one (where z < 0), and we have indicated these signs by plus and minus signs in Fig. 12.14.

Let us consider first an atom with just one unpaired p electron; for example, fluorine, with configuration

$$\text{F: } 1s^2 2s^2 2p_x^2 2p_y^2 2p_z^1$$

The first eight electrons in F are all paired and can, for our present purposes, be ignored. The single unpaired electron, which is distributed as in Fig. 12.14(a), can pair with the one electron in a hydrogen atom to form the molecule HF. To arrange this, however, we must bring up the H atom in such a way that the wave functions of the shared electrons interfere constructively in their region of overlap; this requires that the H atom approach the positive lobe of the F wave function as in Fig. 12.15(a). If the H atom approaches from the opposite direction, as shown in Fig. 12.15(b), we obtain an antibonding orbital, and no molecule forms.

We can summarize our findings so far: When valence electrons are in s states, their distribution is spherical, with no preferred direction for bonding. When the valence electrons are in p states, their distribution is concentrated along a definite axis and this means that they form directed bonds.

We are now ready to discuss some molecules that are more interesting, such as the water molecule, H₂O. The oxygen atom, with configuration

$$\text{O: } 1s^2 2s^2 2p_x^2 2p_y^1 2p_z^1$$

has two unpaired electrons, whose distributions are sketched in Fig. 12.16(a). Each of the unpaired electrons can form a bond with one H atom, but only if the H atoms approach along the y and z axes, as shown in Fig. 12.16(b). This implies that the water molecule should be shaped like a letter "L," as shown schematically in Fig. 12.16(c). It is found that the water molecule does have an "L" shape, but the measured angle between the two arms is about 105°, some 15° larger than predicted. However, the reason for this discrepancy is not hard to find: The two H atoms, having shifted their electrons toward the oxygen, are left somewhat positive and repel one another, pushing the two arms a little farther apart than the 90° predicted by our simple model.

If the bonds in H₂O were purely covalent, the shared electrons would concentrate at the midpoint between the O and each H. It is found, however, that they concentrate somewhat closer to the oxygen. In other words, the bonds are partly ionic and there is a small net negative charge near the O, leaving a small excess of positive charge near the two hydrogens, as shown schematically in

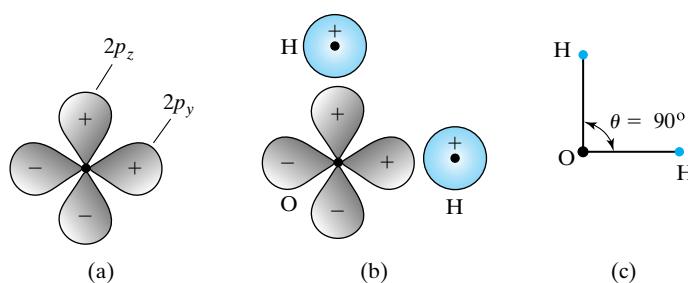


FIGURE 12.16

(a) The two unpaired electrons in an O atom occupy the 2p_y and 2p_z orbitals. (b) A water molecule can form if two H atoms approach the positive lobes of these two orbitals. (c) The centers of the atoms in the resulting molecule form an "L" with angle $\theta = 90^\circ$ between the arms. (The experimental value is $\theta \approx 105^\circ$, as explained in the text.)

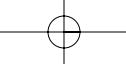
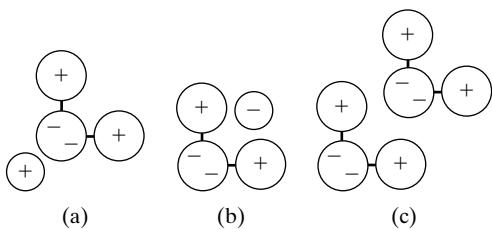
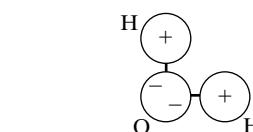
Section 12.6 • Excited States of Molecules **389**

Fig. 12.17. For this reason the H_2O molecule has an electric dipole moment, which is responsible for several familiar properties of water: A water molecule can easily pick up a charge of either sign, as shown in Fig. 12.18(a) and (b), and can carry away charges from a body that is electrostatically charged. This explains why walking across a nylon carpet on a moist day does not cause the annoying electric shocks that we experience when the humidity is low.

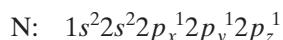
For much the same reason, water molecules can dissolve ionic molecules such as NaCl by attaching themselves to the separate Na^+ and Cl^- ions. Finally, when properly oriented, two water molecules attract each other, as in Fig. 12.18(c). This *dipole-dipole* attraction is what holds water together in snow and ice crystals.

**FIGURE 12.17****FIGURE 12.17**

Schematic view of the charge distribution in a water molecule. The plus and minus signs here indicate positive and negative charges whose observed magnitude is about $e/3$ (Problem 12.32).

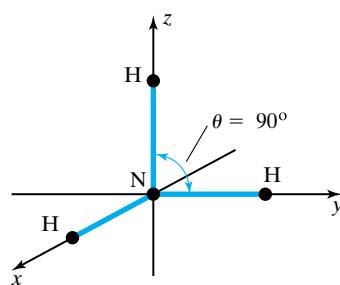
FIGURE 12.18
(a) and **(b)** A water molecule can attach itself to a positive or negative charge. **(c)** Water molecules tend to align and attract one another.

Just as with water, H_2O , we can predict the shape of the ammonia molecule, NH_3 . The nitrogen atom, with configuration



has three unpaired p electrons, each concentrated near one of three perpendicular axes. Thus N can bond with three atoms to form NH_3 , with one H atom on each of the three axes, as shown in Fig. 12.19. According to our simple model, the angle θ between any two of the N—H bonds should be $\theta = 90^\circ$, but, just as with water, the repulsion between the positive protons increases θ somewhat ($\theta = 107^\circ$ in this case).

Similar considerations apply to many other molecules. For example, it can be shown that in certain carbon compounds the four valence electrons of the carbon atom have lobes that point symmetrically outward toward the corners of a regular tetrahedron centered on the nucleus.* This has important implications for the shapes of the many hydrocarbon molecules mentioned at the end of the preceding section. For example, the methane molecule, CH_4 , is a regular tetrahedron with the four H atoms at its corners and the C atom at its center. Since the basic ideas remain the same, while the details get much more complicated, we will not pursue this topic further.

**FIGURE 12.19**

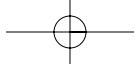
The ammonia molecule, NH_3 , is a pyramid with the N atom at one corner. The H atoms at the other three corners form an equilateral triangle. The angle θ between any two N—H bonds is actually a little more than 90° because of the Coulomb repulsion between the H nuclei.

12.6 Excited States of Molecules*

*This section can be omitted without serious loss of continuity, though you should read it if you want to study Section 12.7.

We have so far discussed only the ground states of molecules. However, most information about molecules comes from the study of molecular spectra; and since spectra result from transitions, which necessarily involve at least one excited state, we now extend our discussion to include excited states as well.

*This arrangement occurs in diamond and methane (CH_4), for example. The valence electrons occupy four orbitals called *s-p* hybrids. Each of these is a combination of the usual *s* and *p* wave functions, and each concentrates the electron in a single lobe. The electrostatic repulsion of the electrons pushes these as far apart as possible, so that they point symmetrically outward as described.



390 Chapter 12 • Molecules

Electronic Levels

Our analysis of molecular ground states started by considering two atoms, at large separation R , both in their ground states with total energy $E(R)$. We found that as the atoms move closer, the curve $E(R)$ may decrease and go through a minimum. If this happens (as it does with H_2 , for example), the two atoms can bond to form a stable molecule, whose ground-state energy is simply the minimum value of $E(R)$.

Suppose now that we start instead with one or more of the atom's valence electrons* in excited states, so that the total energy of the two atoms is $E'(R) > E(R)$. As the two atoms approach one another, it may happen that the curve $E'(R)$ goes through a minimum, and if this does occur, we conclude that this molecule has an excited state with energy equal to the minimum value of $E'(R)$. We can describe this excited state as "stable" in the sense that the atoms will not fly apart spontaneously; however, it is not perfectly stable since it can drop to the ground state (or any other lower level) by emitting radiation. In Fig. 12.20 we have sketched three molecular energy levels (labeled 0, 1, 2) that come about in this way.

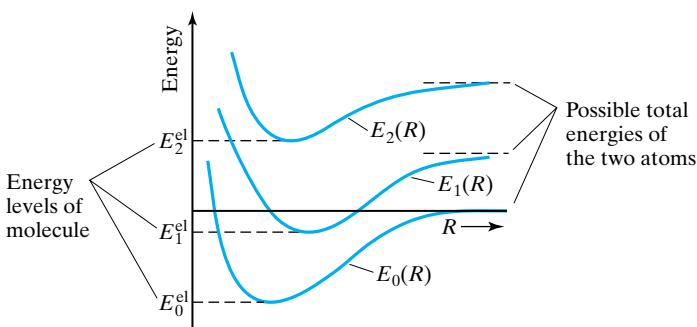
The three curves in Fig. 12.20 give the possible energies of the molecule as functions of the separation R . When R is large, the energies $E_0(R)$, $E_1(R)$, ... are the allowed total energies of the two separated atoms. The minimum values of the functions $E_0(R)$, $E_1(R)$, ... are the corresponding energy levels of the molecule in question. These levels are called **electronic levels** because they are the energies calculated using the Schrödinger equation for the electrons while ignoring possible motions of the two nuclei; for this reason, they are denoted by E_0^{el} , E_1^{el} , ..., as shown on the left of Fig. 12.20. As the picture suggests, one would expect the spacing of these electronic levels of the molecule to be of the same order as the spacing of the atomic levels from which they evolved. Therefore, since the spacing of valence levels in atoms is usually a few eV, the same is generally true for the electronic levels in molecules. This means that transitions between electronic levels in a molecule involve photons of a few eV — that is, photons in or near the visible range.

Motion of the Nuclei

If the electronic levels E_0^{el} , E_1^{el} , ... just described were the only levels of molecules, molecular spectra would look very like atomic spectra. There is, however, a very important complication that we have so far ignored. Up to now,

FIGURE 12.20

When two atoms are far apart, their total energy can have various different values $E_0(R)$, $E_1(R)$, $E_2(R)$, ... As R decreases, each of the curves $E_i(R)$ ($i = 0, 1, 2, \dots$) may have a minimum, and if this happens, there is a corresponding bound state of the molecule. The resulting energy levels of the molecule are denoted by E_i^{el} , as explained in the text.



*Our only concern here is with valence electrons, since inner-shell electrons play little part in molecular bonding.

Section 12.6 • Excited States of Molecules **391**

we have assumed that the two nuclei of our diatomic molecule are at rest; that is, we have considered only the motion of the electrons and have completely ignored the nuclear motion. (Note that by “nuclear motion” we do *not* mean the internal motion of the nuclear constituents, but rather the motion of each nucleus as a whole.) We now consider the energy E^{nuc} associated with this motion and will find that the values of E^{nuc} are quantized, with allowed values $E_0^{\text{nuc}}, E_1^{\text{nuc}}, \dots$. Thus, the allowed energies of the entire molecule, $E = E^{\text{el}} + E^{\text{nuc}}$, are

$$E = E_i^{\text{el}} + E_j^{\text{nuc}} \quad (i = 0, 1, 2, \dots; j = 0, 1, 2, \dots)$$

When the molecule makes a transition, both the states of the electrons (labeled by i) *and* the nuclei (labeled by j) can change. If we consider a transition (up or down) between a level E and a higher level E' ,

$$E = E_i^{\text{el}} + E_j^{\text{nuc}} \leftrightarrow E' = E_{i'}^{\text{el}} + E_{j'}^{\text{nuc}}$$

then the photon absorbed or emitted in the transition has energy

$$\begin{aligned} E_\gamma &= E' - E = (E_{i'}^{\text{el}} - E_i^{\text{el}}) + (E_{j'}^{\text{nuc}} - E_j^{\text{nuc}}) \\ &= \Delta E^{\text{el}} + \Delta E^{\text{nuc}} \end{aligned} \quad (12.22)$$

The energies ΔE^{el} are, as we just argued, typically a few eV. On the other hand, the energies ΔE^{nuc} are generally small fractions of an eV and are small corrections to ΔE^{el} . However, there are many possible different values for ΔE^{nuc} . Therefore, (12.22) implies that the spectrum of photons absorbed or emitted, instead of being a single line for each value of ΔE^{el} , will consist of many different lines, all close to ΔE^{el} , corresponding to the many possible small values of ΔE^{nuc} . Unless a spectrometer has high resolution, it may fail to resolve some of these lines, which will then appear instead as a wide **band**. Several such bands in the emission spectrum of N_2 gas are illustrated in Fig. 12.21. In the low-resolution picture one sees several broad lines or bands; in the high-resolution picture three of these bands are seen clearly to be made of many narrowly spaced lines. The occurrence of such **band spectra** at visible, or nearly visible, frequencies is one of the chief characteristics of molecular spectra, and the measurement of their many constituent frequencies is a rich source of information about molecules.

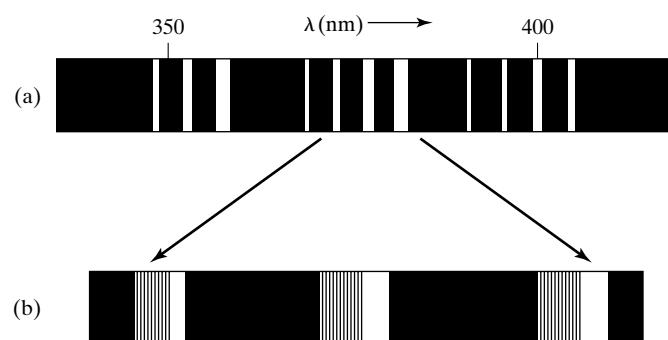
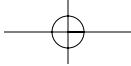


FIGURE 12.21

Emission spectrum of N₂ gas in the near ultraviolet. (a) Several broad lines, or bands, between 340 and 410 nm. (b) High-resolution enlargement of the three bands near 380 nm shows that each consists of many separate lines.



392 Chapter 12 • Molecules

Vibrational Energy

For a diatomic molecule, there are two kinds of motion that contribute to E^{nuc} . First, the two nuclei can move in and out along the internuclear axis, with a **vibrational energy** E^{vib} ; and second, they can rotate about their center of mass with a **rotational energy** E^{rot} . Thus

$$E^{\text{nuc}} = E^{\text{vib}} + E^{\text{rot}}$$

We will consider these two contributions in turn.

The simplest calculation of the vibrational energy of the nuclei uses an approach first proposed by Max Born and the American physicist Robert Oppenheimer. Because the electrons are so much lighter, they move much faster than the nuclei. Thus, the electrons will complete many orbits before the nuclei move appreciably. This means that the electron orbits can be calculated as if for fixed nuclei and that the electrons will occupy these “fixed-nuclei” orbits, adjusting themselves continuously to follow the slow changing of the nuclear positions. (Although this argument is phrased classically, the same conclusion holds in quantum mechanics as well.) This result has two important consequences: First, our calculation of the electronic energies $E_i(R)$, based on stationary nuclei, is not wasted. As the nuclear separation R varies slowly, the electron orbits adjust themselves and, for each value of R , the energy is still given correctly by $E_i(R)$.

The second important consequence of the Born–Oppenheimer approach is that it shows us how to find the vibrational energies. As the nuclei move slowly in and out, they carry the electrons with them, and we can just as well think of the vibrations as vibrations of the two whole atoms, as illustrated in Fig. 12.22. When the separation of the atoms is R , their total energy is the energy $E_i(R)$, already discussed, *plus* the kinetic energy of the vibrating atoms:

$$E_{\text{tot}} = K + E_i(R) \quad (12.23)$$

Here K denotes the kinetic energy of the two atoms each moving as a whole, and this is the term that we have hitherto ignored. Let us imagine for a moment that our two atoms are classical particles with potential energy $U(R)$ and total energy

$$E_{\text{tot}} = K + U(R) \quad (12.24)$$

Since total energy is conserved, the variations of K must exactly mirror those of $U(R)$ — when $U(R)$ increases, K must decrease equally, and vice versa. That is, $U(R)$ governs the motions of the two atoms. Returning to (12.23), we see that since E_{tot} is conserved, the variations of K must similarly mirror those of $E_i(R)$. That is, $E_i(R)$ in (12.23) plays the role of the potential energy that governs the motion of the two atoms. Although this discussion is classical, the corresponding result for a quantum system is that $E_i(R)$ in (12.23) is the effective potential-energy function in the Schrödinger equation for the two atoms. To find the allowed vibrational energies of our molecule, we must solve the Schrödinger equation for two bodies (the two atoms) with the potential-energy function $U(R) = E_i(R)$.

In Figure 12.20 we sketched three of the functions $E_i(R)$; what we have now established is that each of these curves defines a potential well, which controls the vibrational motion of the two atoms. The different curves $E_0(R)$,

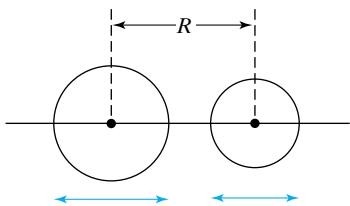
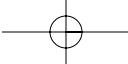
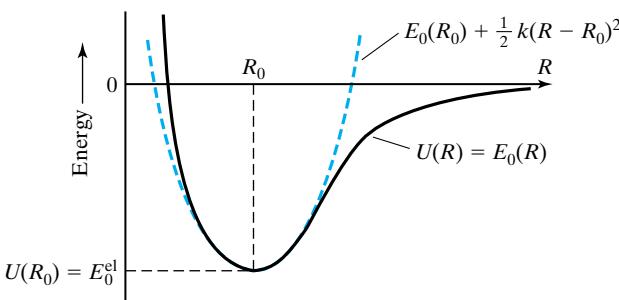


FIGURE 12.22

The vibrational motion of a diatomic molecule is an in-and-out motion of the two atoms along the internuclear axis. (In reality, the two atoms overlap one another, but for clarity we have drawn them separately.)

Section 12.6 • Excited States of Molecules **393****FIGURE 12.23**

If the electrons in a molecule occupy their ground state with energy $E_0(R)$, the potential-energy function that controls the slow vibrations of the two atoms is $U(R) = E_0(R)$. Provided that the nuclei remain close to R_0 the equilibrium separation, $U(R)$ can be approximated by a parabola as shown by the dashed curve.

$E_1(R), \dots$ correspond to different states of the electrons, and we see that for each electronic state, $i = 0, 1, \dots$, there is a different potential-energy function, $U(R) = E_i(R)$, to govern the vibrational motion.

Let us focus attention first on the electronic ground state, $i = 0$, with potential-energy function $U(R) = E_0(R)$. To simplify our discussion, we consider a molecule, such as HCl, in which one atom is much lighter than the other. In this case we can assume, to a good approximation, that the heavy atom (mass m') is stationary, and we have only to consider the motion of the light atom (mass $m \ll m'$).^{*} As a further simplification, we note that the vibrational motion occurs along the line joining the two nuclei. Thus our problem is reduced to solving the Schrödinger equation for a single body (the lighter atom) moving in one dimension, with potential energy $U(R) = E_0(R)$, as sketched in Fig. 12.23.

As long as the atoms remain reasonably close to their equilibrium separation R_0 , the well of Fig. 12.23 can be closely approximated by a parabola of the form

$$E_0(R) \approx E_0(R_0) + \frac{1}{2}k(R - R_0)^2 \quad (12.25)$$

for some suitably chosen constant k , as shown by the dashed curve in the picture (see Problems 12.37 and 12.43). This potential energy is the simple harmonic oscillator, or SHO, potential, discussed in Section 7.9, where k was called the force constant. We found there that the allowed energies are [see Eq. (7.98)][†]

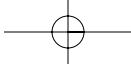
$$E_n^{\text{vib}} = (n + \frac{1}{2})\hbar\omega_c \quad (n = 0, 1, 2, \dots) \quad (12.26)$$

where ω_c is the angular frequency of the corresponding classical oscillator (same mass m and force constant k):

$$\omega_c = \sqrt{\frac{k}{m}}$$

*Actually, the case that m and m' are of the same order is not much more difficult. As described briefly in Section 5.8 and Problem 5.21, the two-body problem with both masses moving is equivalent to the problem with one of the bodies fixed, except that one must use the reduced mass, $\mu = mm'/(m + m')$ in place of m . Thus all of the formulas of this section are correct, whether or not $m \ll m'$, provided that one replaces m by μ . Note that if m is much less than m' , then $\mu \approx m$ (see Problems 12.39 and 12.40).

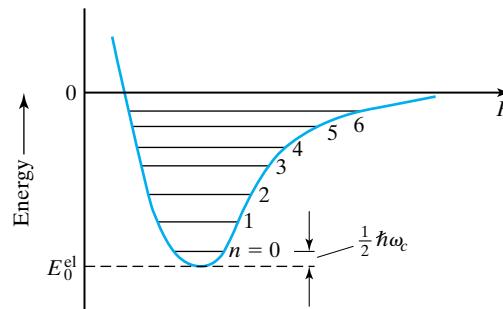
†Equation (12.26) gives the allowed energies measured up from the bottom of the well (12.25), and the total energy is therefore $E_0(R_0) + (n + \frac{1}{2})\hbar\omega_c$. The first term $E_0(R_0)$ is the electronic energy E_0^{el} , and the term $(n + \frac{1}{2})\hbar\omega_c$ is the energy of vibration, which we denote by E_n^{vib} .



394 Chapter 12 • Molecules

FIGURE 12.24

Vibrational levels of a diatomic molecule in its lowest electronic state. The numbers on the right identify the vibrational quantum number, n . For clarity, we have shown a molecule with only seven levels; in most real molecules there are more vibrational levels, more closely spaced.



The lowest vibrational level has the zero-point energy $\frac{1}{2}\hbar\omega_c$, and all the levels given by (12.26) are spaced at equal intervals $\hbar\omega_c$. However, at high energies the approximation (12.25) breaks down and, as seen in Fig. 12.23, the true potential well rapidly becomes wider than its parabolic approximation. This means that the higher levels are somewhat closer together, and there are no levels above $E = 0$. This is illustrated in Fig. 12.24.

We see in Fig. 12.24 that since the lowest possible energy is $E_0^{\text{el}} + \frac{1}{2}\hbar\omega_c$, the ground state of the molecule is higher by $\frac{1}{2}\hbar\omega_c$ than the value E_0^{el} we used in Sections 12.3 and 12.4 (where we ignored motion of the nuclei). However, this zero-point energy is usually quite small compared to electronic energies. For example, in LiBr, it is about 0.035 eV, as we see in the following example. Compared to typical electronic energies of order 1 eV, this is indeed fairly small.

Example 12.4

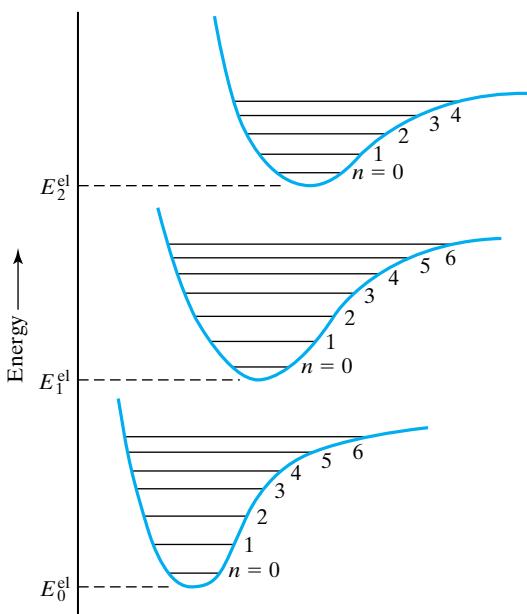
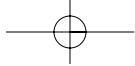
The parabola that best approximates the potential well $E_0(R)$ for LiBr has a force constant $k \approx 800 \text{ eV/nm}^2$ (see Problem 12.47). Use this to find the zero-point energy $\frac{1}{2}\hbar\omega_c$ for LiBr.

The mass of Br (80 u) is much greater than that of Li (7 u), so we can make the approximation that the Br is fixed, and we get

$$\begin{aligned}\frac{1}{2}\hbar\omega_c &= \frac{\hbar}{2}\sqrt{\frac{k}{m}} = \frac{\hbar c}{2}\sqrt{\frac{k}{mc^2}} \\ &= \frac{200 \text{ eV} \cdot \text{nm}}{2} \sqrt{\frac{800 \text{ eV}/\text{nm}^2}{7 \times 931.5 \times 10^6 \text{ eV}}} \approx 0.035 \text{ eV}\end{aligned}$$

Note that the spacing between adjacent vibrational levels is just twice this, $\hbar\omega_c \approx 0.07 \text{ eV}$, which is also small compared to typical electronic energies.

We can apply similar considerations to any other electronic state. For each electronic state, labeled i , the curve $E_i(R)$ defines the oscillator well that controls the vibrational motion of the atoms, and for each such well there will be a ladder of vibrational levels $E_n^{\text{vib}} = (n + \frac{1}{2})\hbar\omega_c$. The levels of the molecule, including electronic and vibrational motion, are therefore as sketched in Fig. 12.25.

**FIGURE 12.25**

Energy levels of a typical diatomic molecule, including electronic and vibrational energies. The three electronic levels are shown as E_0^{el} , E_1^{el} , E_2^{el} . For each electronic level, there is a ladder of vibrational energies, $E_n^{\text{vib}} \approx (n + \frac{1}{2})\hbar\omega_c$ with $n = 0, 1, \dots$; these are drawn inside the appropriate well, with the quantum number n shown at the right.

Rotational Energy

The rotational energies are easily calculated. We can think of the molecule as a dumbbell, which can rotate about its center of mass as shown in Fig. 12.26. The classical energy of rotation is $E^{\text{rot}} = \frac{1}{2}I\omega^2$, where I is the moment of inertia and ω is the angular velocity. Since the angular momentum L is $I\omega$, it follows that $\omega = L/I$ and the rotational energy is*

$$E^{\text{rot}} = \frac{L^2}{2I}$$

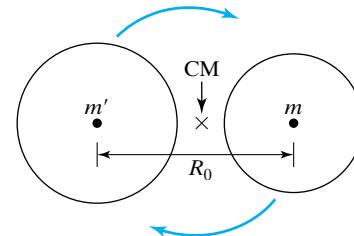
This formula is also valid in quantum mechanics, provided that we use the correct quantized values $l(l+1)\hbar^2$ for L^2 . Hence, the allowed values of the rotational energy of our diatomic molecule are

$$E_l^{\text{rot}} = l(l+1)\frac{\hbar^2}{2I} \quad (l = 0, 1, 2, \dots) \quad (12.27)$$

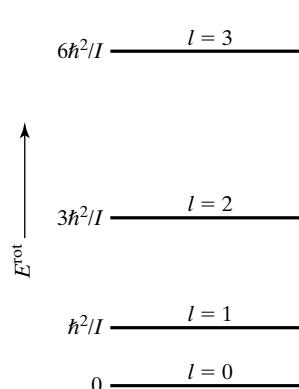
These values of E^{rot} are shown in Fig. 12.27, where we see — what is also clear from Eq. (12.27) — that the energies are all positive and that the spacing between the levels increases steadily with increasing l .

To estimate the magnitude of the rotational energies (12.27), we again consider a molecule like HCl or LiBr in which one atom is much lighter than the other. In this case the rotational motion is — to a good approximation — just an orbiting of the light atom (mass m') around the fixed heavy atom, and the moment of inertia is simply

$$I = mR_0^2$$

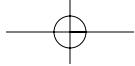
**FIGURE 12.26**

Rotational motion of a diatomic molecule.

**FIGURE 12.27**

Quantized energies of rotation, $l(l+1)\hbar^2/2I$, for a diatomic molecule.

*Notice that this result is the rotational analog of the more familiar result for the translational energy $p^2/2m$.



396 Chapter 12 • Molecules

In the case of HCl the observed value of R_0 is 0.13 nm and the light atom (H) has mass $m = 938 \text{ MeV}/c^2$. Thus, for example, the $l = 1$ level has energy

$$\begin{aligned}\frac{\hbar^2}{I} &= \frac{\hbar^2}{mR_0^2} = \frac{(\hbar c)^2}{mc^2 R_0^2} \\ &\approx \frac{(200 \text{ eV} \cdot \text{nm})^2}{(938 \times 10^6 \text{ eV}) \times (0.13 \text{ nm})^2} \approx 2 \times 10^{-3} \text{ eV}. \quad (12.28)\end{aligned}$$

Since the electronic energies are of order 1 eV while the energies of vibration are of order 0.1 eV, we see that rotational energies are much smaller than both electronic and vibrational energies.

Total Energy of the Molecule

We can now put our various results together. The total energy of a diatomic molecule is

$$E = E^{\text{el}} + E^{\text{vib}} + E^{\text{rot}} \quad (12.29)$$

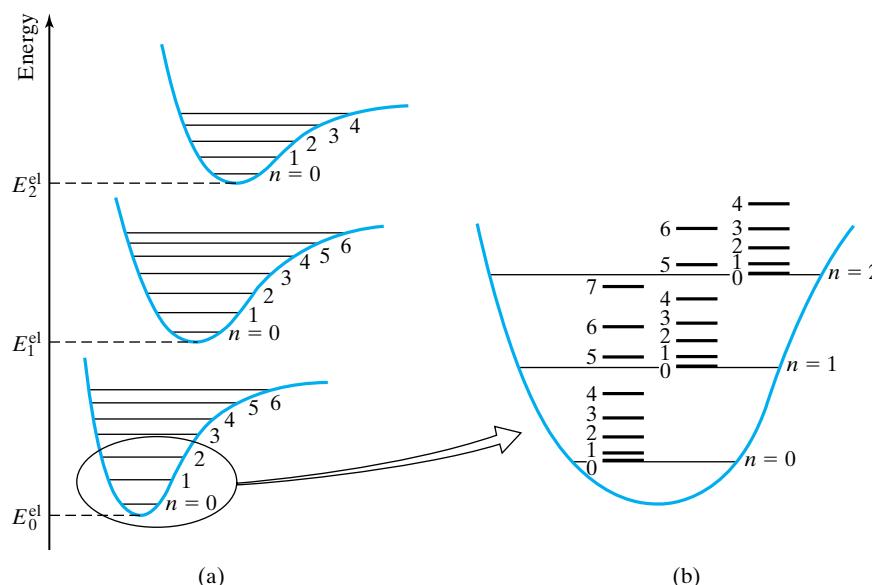
All three of these energies are quantized. The allowed values of E^{el} we have denoted by $E_i^{\text{el}} (i = 0, 1, \dots)$; those for E^{vib} are denoted $E_n^{\text{vib}} (n = 0, 1, \dots)$ and, at least for the lower levels, have the approximate form $(n + \frac{1}{2})\hbar\omega_c$; finally, the allowed values of E^{rot} are $l(l + 1)\hbar^2/2I$. Representative values for the spacings of the various levels are

$$\begin{aligned}\Delta E^{\text{el}} &\sim 1 \text{ eV} \\ \Delta E^{\text{vib}} &\sim 0.1 \text{ eV} \\ \Delta E^{\text{rot}} &\sim 0.001 \text{ eV}\end{aligned}$$

If we ignore the small rotational energies entirely, the energy levels of the molecule appear as shown in Fig. 12.25 [which we have reproduced as Fig. 12.28(a)]. Each electronic state, $i = 0, 1, \dots$, defines a potential well that

FIGURE 12.28

Energy levels of a diatomic molecule have the form $E_n^{\text{el}} + E_n^{\text{vib}} + E_l^{\text{rot}}$ (a) For each electronic level, there is a ladder of vibrational levels, labeled $n = 0, 1, \dots$ (On the scale of this picture the rotational levels are too closely spaced to be shown.) (b) Above each vibrational level is a "subladder" of very closely spaced rotational levels, as illustrated by this enlargement of the three lowest vibrational levels. The rotational levels are labeled by the quantum number $l = 0, 1, 2, \dots$, shown on their left. Note that in most molecules the rotational levels are much more closely spaced than shown here.



determines the vibrational motion of the atoms, and inside each well we have drawn the ladder whose rungs represent the different vibrational energies $E_n^{\text{vib}} \approx (n + \frac{1}{2})\hbar\omega_c$, with $n = 0, 1, \dots$.

If we wish to take into account the rotational energies, then above each rung of every ladder in Fig. 12.28(a) we must draw a “subladder” whose rungs represent the rotational energies $l(l+1)\hbar^2/2I$ ($l = 0, 1, \dots$). On the scale of Fig. 12.28(a) the rungs of these subladders are too closely spaced to be distinguished, but Fig. 12.28(b) shows an enlargement of the lowest three levels of part (a), each with its subladder of rotational levels. It is the level structure in Fig. 12.28 — with rotational levels built on vibrational levels built on electronic levels — that determines the main features of molecular spectra, as we describe in the next section.

12.7 Molecular Spectra*

*This section can be omitted without significant loss of continuity.

Knowing the structure of molecular energy levels — with their hierarchy of electronic, vibrational, and rotational energies — we can now discuss the spectrum of radiation emitted or absorbed as a molecule changes its energy. The many possible different spectral lines can be classified according to which of the terms in the total energy,

$$E = E^{\text{el}} + E^{\text{vib}} + E^{\text{rot}}$$

changes in the corresponding transitions.

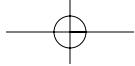
Rotational Transitions

We begin by discussing the **rotational transitions**, in which only the rotational state changes. Since the separation of rotational levels is of order 10^{-3} eV, the same is true of the photons emitted or absorbed in rotational transitions. Photons with energy 10^{-3} eV have wavelength of order 1 mm and are classified as *microwave* photons. Thus, the rotational spectrum of most molecules lies in the microwave region. Microwave ovens exploit this fact; the frequency of the radiation in these ovens is designed to match the rotational modes of the water molecule.

At room temperature and below, the average kinetic energy of the molecules in a gas is of order 0.04 eV or less.* This is so low that collisions between molecules will seldom excite either the electronic or vibrational levels. On the other hand, it is quite enough to excite several rotational levels. Thus the molecules of a gas at room temperature remain locked in their lowest electronic and vibrational levels, but are continually being excited to several different rotational levels. In many molecules transitions between different rotational levels can occur rapidly[†] and the gas will emit microwave radiation as these excited molecules drop back to lower levels. Since this radiation is rather weak, the most convenient way to study the rotational spectrum is usually to

*We know from the kinetic theory of gases (Sec. 3.7) that the average translational kinetic energy is $3k_B T/2$, where k_B denotes Boltzmann's constant, $k_B = 8.62 \times 10^{-5}$ eV/K. At room temperature ($T \approx 300$ K) this energy is about 0.04 eV.

[†]The main exception is that in homonuclear molecules (molecules such as H₂ with two identical nuclei) a selection rule forbids these transitions and there is no rotational spectrum.



398 Chapter 12 • Molecules

monitor the absorption of a microwave beam that is directed through the gas and whose frequency can be adjusted. The different transitions then reveal themselves as sharp reductions in the transmitted beam whenever the microwave photons have the right energy to excite the molecule from one level to another.

Not all transitions that are energetically possible occur with appreciable probability. As we discussed in Section 11.8, those transitions that have appreciable probability are called *allowed* and are determined by *selection rules*. The selection rule for transitions between different rotational levels is*

$$\Delta l = \pm 1 \quad (12.30)$$

That is, only those transitions in which l changes by one unit occur rapidly enough to be easily observed. Therefore, the only allowed rotational transitions, up or down, have the form

$$l \leftrightarrow l + 1 \quad (l = 0, 1, 2, 3, \dots) \quad (12.31)$$

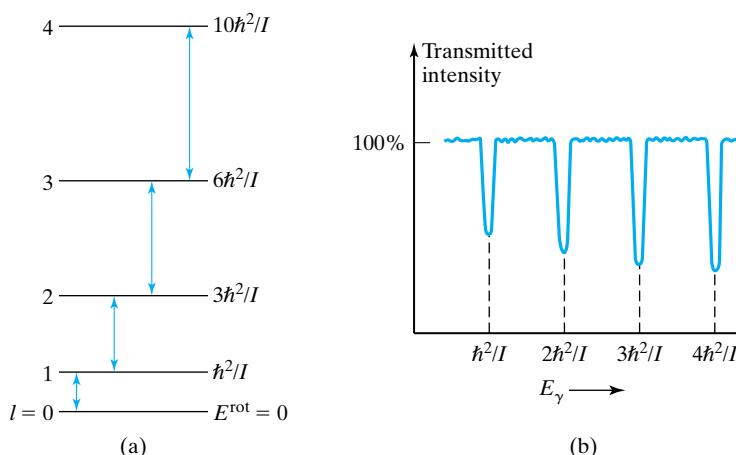
as shown in Fig. 12.29(a). Since the energy E_l^{rot} of the rotational levels is $l(l+1)\hbar^2/2I$, the energies of the photons emitted or absorbed in these transitions have the form

$$\begin{aligned} E_{\gamma}(l \leftrightarrow l + 1) &= E_{l+1}^{\text{rot}} - E_l^{\text{rot}} = [(l+1)(l+2) - l(l+1)] \frac{\hbar^2}{2I} \\ &= \frac{(l+1)\hbar^2}{I} \quad (l = 0, 1, 2, 3, \dots) \end{aligned} \quad (12.32)$$

This gives a series of equally spaced lines separated by the energy \hbar^2/I , as illustrated in Fig. 12.29(b). By measuring the separation of any two adjacent lines in the rotational spectrum of a molecule, one can determine the molecule's moment of inertia, I , and its bond length R_0 .

FIGURE 12.29

The rotational spectrum of a molecule is produced by transitions between different rotational levels, while the vibrational and electronic levels remain unchanged. (a) The allowed transitions satisfy $\Delta l = \pm 1$. (b) General appearance of the rotational spectrum in absorption. A microwave beam of adjustable frequency is directed through the gas and is absorbed only when its photons have energy ($E = hf$) equal to one of the transition energies $\hbar^2/I, 2\hbar^2/I, \dots$



*This is the same selection rule encountered in (11.46). As mentioned there, it follows because the photon has spin 1 and must carry away exactly one unit of angular momentum.

Example 12.5

The four absorption lines of longest wavelength in lithium iodide gas are found at $\lambda = 1.13, 0.564, 0.376$, and 0.282 cm . Deduce the bond length of the LiI molecule.

Because the lines with longest wavelength correspond to photons with lowest energy, the four given lines should correspond to the lowest four transitions of Fig. 12.29. The four energies $E = hc/\lambda$ are easily calculated.

$\lambda \text{ (cm)}$	$E_\gamma \text{ (cm)}$	Transition
1.13	1.10×10^{-4}	$0 \rightarrow 1$
0.564	2.20×10^{-4}	$1 \rightarrow 2$
0.376	3.30×10^{-4}	$2 \rightarrow 3$
0.282	4.40×10^{-4}	$3 \rightarrow 4$

These are seen to be in the ratios $1:2:3:4$, as required by (12.32), with

$$\frac{\hbar^2}{I} = \frac{\hbar^2}{mR_0^2} = 1.10 \times 10^{-4} \text{ eV}$$

Since lithium has $m = 7\text{u} = 7 \times 931.5 \text{ MeV}/c^2$, this gives

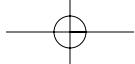
$$\begin{aligned} R_0 &= \frac{\hbar}{\sqrt{m(1.10 \times 10^{-4} \text{ eV})}} = \frac{\hbar c}{\sqrt{mc^2(1.10 \times 10^{-4} \text{ eV})}} \\ &= \frac{197 \text{ eV} \cdot \text{nm}}{\sqrt{(7 \times 931.5 \times 10^6 \text{ eV}) \times (1.10 \times 10^{-4} \text{ eV})}} = 0.233 \text{ nm} \end{aligned}$$

Note that in this calculation we treated the iodine nucleus (mass = 127 u) as fixed, so that $I = mR_0^2$ with $m = 7 \text{ u}$, the mass of the lithium nucleus. If one takes account of the small motion of the iodine, one gets $R_0 = 0.239 \text{ nm}$ (see Problem 12.49).

Rotational spectra of molecules are useful in astronomy, where they make possible the identification of many molecules in the large “molecular clouds” of our galaxy. These clouds are often so cool (a few tens of degrees kelvin) that intermolecular collisions almost never raise the molecules to excited vibrational or electronic levels; nevertheless, they can excite several rotational levels, which then emit radiation that can be detected by radio telescopes on earth. For example, carbon monoxide is frequently detected in molecular clouds by the microwaves emitted in the transitions $l = 1 \rightarrow 0$ and $l = 2 \rightarrow 1$.

Vibrational–Rotational Transitions

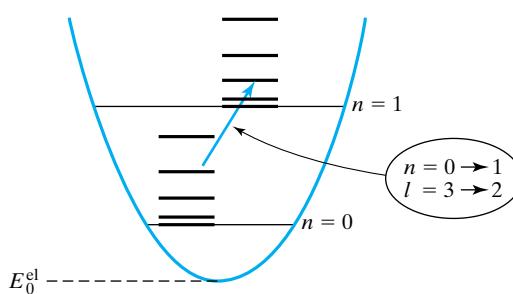
In many molecules one can observe transitions in which the vibrational level changes but the electronic level does not. However, in any such transition the selection rule (12.30) still applies, and the rotational level must also change. For this reason, these transitions are called **vibrational–rotational transitions**. We have seen that the spacing of vibrational levels is of order 0.1 eV, while that of rotational levels is much smaller. Therefore, the photons emitted or absorbed in vibrational–rotational transitions have energies of order 0.1 eV and are in the infrared.



400 Chapter 12 • Molecules

FIGURE 12.30

In a vibrational-rotational transition the electronic state is unchanged, but both n and l change by one unit. In the transition shown, n increases ($n = 0 \rightarrow 1$) and l decreases ($l = 3 \rightarrow 2$). Since the total energy increases, this transition entails the absorption of a photon. (Spacing of the rotational levels is greatly exaggerated.)



In addition to (12.30) there is also a selection rule for the vibrational quantum number.

$$\Delta n = \pm 1 \quad (12.33)$$

That is, among the many conceivable vibrational-rotational transitions, only those in which both n and l change by one unit are observed with appreciable probability. A typical such transition is illustrated in Fig. 12.30. The energies involved have the form

$$E_n^{\text{vib}} = \left(n + \frac{1}{2}\right)\hbar\omega_c \quad \text{and} \quad E_l^{\text{rot}} = \frac{l(l+1)\hbar^2}{2I} \quad (12.34)$$

with the vibrational spacing ($\hbar\omega_c$) much greater than the rotational. Therefore, if n decreases by one ($n \rightarrow n - 1$), the total energy decreases whichever way l changes ($l \leftrightarrow l + 1$), and a photon is emitted with energy

$$\begin{aligned} E_\gamma &= (E_n^{\text{vib}} - E_{n-1}^{\text{vib}}) \pm (E_{l+1}^{\text{rot}} - E_l^{\text{rot}}) \\ &= \hbar\omega_c \pm \frac{(l+1)\hbar^2}{I} \quad (l = 0, 1, 2, \dots) \end{aligned} \quad (12.35)$$

If n increases by one, the molecule's energy increases and a photon has to be absorbed, with energy again given by (12.35). Either way, (12.35) implies that the vibrational-rotational spectrum consists of many lines, equally spaced on either side of $\hbar\omega_c$, with spacing \hbar^2/I . Since there are no transitions with $\Delta l = 0$, there is no line with E_γ actually equal to $\hbar\omega_c$.

A typical vibrational-rotational absorption spectrum, for HBr, is shown in Fig. 12.31. One sees clearly the equally spaced lines as predicted by (12.35),

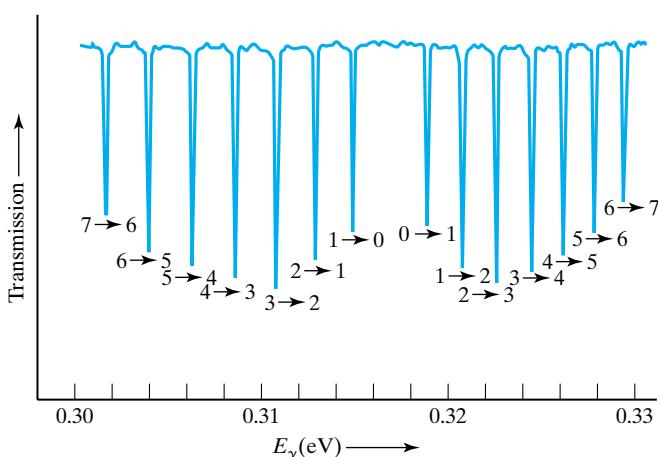
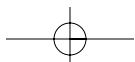


FIGURE 12.31

Vibrational-rotational absorption spectrum of HBr. All lines correspond to the vibrational transition $n = 0 \rightarrow 1$, with various different rotational transitions $l \leftrightarrow l - 1$, as indicated.



with a gap at the center corresponding to the absence of transitions with $E_\gamma = \hbar\omega_c$. From the position of the central gap, one can measure the vibrational parameter $\hbar\omega_c$; and from the spacing of the lines, one can read off the rotational parameter \hbar^2/I , as the following example shows.

Example 12.6

Use the graph in Fig. 12.31 to determine the parameters $\hbar\omega_c$ and \hbar^2/I for the HBr molecule. Hence, find the force constant k and bond length R_0 for HBr.

The central gap in a vibrational–rotational spectrum comes at $E = \hbar\omega_c$, and from Fig. 12.31, we can read off $\hbar\omega_c$ as

$$\hbar\omega_c = 0.317 \text{ eV} \quad (12.36)$$

Since $\omega_c = \sqrt{k/m}$, where m is the mass of the hydrogen atom (as usual, we ignore the motion of the heavier atom), a simple calculation gives

$$\begin{aligned} k = m\omega_c^2 &= \frac{mc^2(\hbar\omega_c)^2}{(\hbar c)^2} = \frac{(938 \times 10^6 \text{ eV}) \times (0.317 \text{ eV})^2}{(197 \text{ eV} \cdot \text{nm})^2} \\ &= 2.4 \times 10^3 \text{ eV/nm}^2 \end{aligned}$$

The spacing between adjacent lines in a vibrational–rotational spectrum is \hbar^2/I and from Fig. 12.31, we find

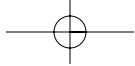
$$\frac{\hbar^2}{I} = 0.0020 \text{ eV} \quad (12.37)$$

Comparing (12.36) and (12.37), we see that the rotational energies are smaller than the vibrational by two orders of magnitude, as stated earlier. Given that $I = mR_0^2$ (where m is the mass of hydrogen), we find R_0 to be

$$\begin{aligned} R_0 &= \sqrt{\frac{I}{m}} = \sqrt{\frac{\hbar^2}{m(\hbar^2/I)}} = \frac{\hbar c}{\sqrt{mc^2(\hbar^2/I)}} \\ &= \frac{197 \text{ eV} \cdot \text{nm}}{\sqrt{(938 \times 10^6 \text{ eV}) \times (0.0020 \text{ eV})}} = 0.14 \text{ nm} \quad (12.38) \end{aligned}$$

The relative strengths of the absorption or emission lines in any spectrum depend on how many of the molecules occupy the initial levels concerned, and this occupancy depends on the temperature of the gas. (The higher the temperature, the greater the number of levels occupied.) Measurement of the relative intensities of the lines in a spectrum like those of Fig. 12.31 or Fig. 12.29(b) allows one to find the temperature of the gas. This is especially useful in astronomy, where the gases of interest are inaccessible to direct measurement.

In fact, rotational and vibrational–rotational spectra are useful to astronomers for several other reasons as well. They can be used to identify the molecules in interstellar clouds that are too cool to emit at higher frequencies. Furthermore, microwave and IR radiations can pass through dust clouds that are completely opaque to visible light. Thus, microwaves and IR can give information about regions of our galaxy that are inaccessible to conventional optical astronomy. For example, they allow direct observations of the core of



402 Chapter 12 • Molecules

the galaxy, where measurements of Doppler shifted spectra let us determine the speed of rotation of the galaxy's inner spiral arms.

Electronic Transitions

Any transition in which the electronic state of a molecule changes is called an **electronic transition**. In electronic transitions both the vibrational and rotational states can also change. A typical electronic transition has the form

$$(E = E_i^{\text{el}} + E_n^{\text{vib}} + E_l^{\text{rot}}) \leftrightarrow (E' = E_i'^{\text{el}} + E_n'^{\text{vib}} + E_l'^{\text{rot}})$$

and the energy of the photon emitted or absorbed is

$$E_\gamma = E' - E = \Delta E^{\text{el}} + \Delta E^{\text{vib}} + \Delta E^{\text{rot}} \quad (12.39)$$

Since the largest of these three terms, ΔE^{el} , is of order a few eV, the same is true of E_γ , and electronic transitions involve photons in the visible, or nearly visible, region.

Just like an atom, a molecule can make both upward and downward transitions, producing absorption and emission spectra. If a gas sample is at room temperature, however, all the molecules will be in the lowest electronic and vibrational states. In this case they can make no downward electronic transitions and will produce no electronic *emission* spectrum. On the other hand, they can make upward transitions — if exposed to light of the right frequency — and their electronic spectrum can be studied in absorption. Any observed transition must start in the lowest electronic and vibrational levels since only these are occupied. If we ignore for a moment the small rotational energies, the observed transitions will have the form

$$(E_0 = E_0^{\text{el}} + E_0^{\text{vib}}) \rightarrow (E = E_i^{\text{el}} + E_n^{\text{vib}}) \quad (12.40)$$

The selection rule $\Delta n = \pm 1$, (12.33), which applies to vibrational transitions in a single SHO well, does not apply to transitions between the levels of two different wells. Therefore, the electronic transitions (12.40) can occur with many different values of n :

$$0 \rightarrow n \quad (n = 0, 1, 2, \dots)$$

as illustrated in Fig. 12.32. For any one final electronic level, these transitions will produce a series of equally spaced absorption lines, corresponding to the several possible final vibrational levels $n = 0, 1, 2, \dots$, as shown in Fig. 12.33(a).

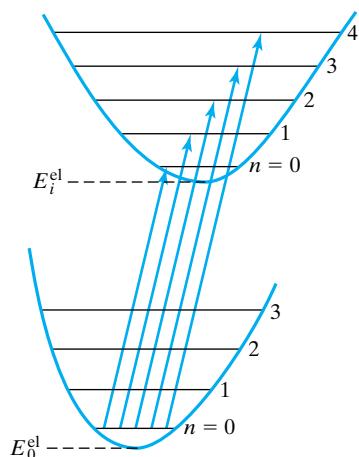


FIGURE 12.32

Typical electronic transitions starting from the lowest electronic and vibrational levels. Since there is no selection rule on n in electronic transitions, many different transitions, $0 \rightarrow n$, ($n = 0, 1, 2, 3, \dots$), are possible for any one final electronic level E_i^{el} .

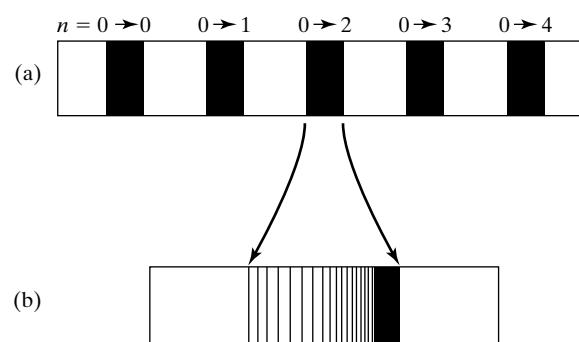


FIGURE 12.33

(a) The electronic absorption spectrum corresponding to the transitions shown in Fig. 12.32.
 (b) At high resolution, each of the bands in part (a) is seen to be a series of lines, corresponding to the many possible changes of rotational energy.

If we take account of the rotational energies, every one of the levels shown in Fig. 12.32 has a ladder of many very closely spaced rotational levels above it. Thus, for each of the transitions (12.40), there are in reality many transitions in which the rotational energy changes by small but different amounts. Thus each of the lines in Fig. 12.33(a) is actually a band and is found, at high enough resolution, to be a series of closely spaced lines, as illustrated in Fig. 12.33(b).

In summary, molecular spectra range from rotational spectra in the microwave region, through vibrational–rotational spectra in the IR, to electronic spectra in the visible region. These provide three different windows, often requiring quite different experimental techniques, through which we study the properties of molecules. They are also the basis of many practical applications, including lasers and masers, as discussed in Chapter 11.

CHECK LIST FOR CHAPTER 12

CONCEPT	DETAILS
Bond length (of diatomic molecule)	Distance between nuclei (Sec. 12.2)
Binding, or dissociation, energy	Energy to pull atoms apart (Sec. 12.2)
Ionic bonds	Atoms exchange one or more electrons and are then bound by the Coulomb attraction (Sec. 12.3). Dipole moments, $p = qd$ (12.1)
Covalent bonds	Atoms held together by shared electrons (Sec. 12.4) H_2^+ and H_2 molecules Bonding and antibonding orbitals Directional properties* (examples of H_2O , NH_3 , and CH_4)
Valence	Number of electrons an atom can gain or lose in an ionic bond. Number of electrons an atom can share in a covalent bond.
Excited states*	$E = E^{\text{el}} + E^{\text{vib}} + E^{\text{rot}}$ (12.29) Electronic levels, $E_i^{\text{el}} \sim 1 \text{ eV}$ Vibrational levels, $E_n^{\text{vib}} = \left(n + \frac{1}{2}\right)\hbar\omega_c \sim 0.1 \text{ eV}$ (12.26) Rotational levels, $E_l^{\text{rot}} = l(l+1)\hbar^2/2I \sim 0.001 \text{ eV}$ (12.27)
Molecular spectra*	Rotational transitions (microwave) (12.32) Vibrational–rotational transitions (IR) (12.35) Electronic transitions (visible or nearly visible) (12.39)

PROBLEMS FOR CHAPTER 12

SECTION 12.2 (Overview of Molecular Properties)

- 12.1** • (a) If all the electrons in 1 gram of hydrogen could be detached from all the protons and separated from them by 1 cm, what would be the total electrostatic force between the electrons and protons? (b) What would be the electrons' acceleration in g's (assuming that they are somehow held together as a single body)?
- 12.2** • If all the electrons in the earth could be removed and attached to the sun (leaving all the protons on

the earth), what would be the resulting electrostatic force of the sun on the earth, and what acceleration would it produce? Compare this force with the corresponding gravitational force. (Make reasonable estimates. The mass of the earth is about $6 \times 10^{24} \text{ kg}$, and that of the sun is $2 \times 10^{30} \text{ kg}$. The distance from earth to sun is $1.5 \times 10^{11} \text{ m}$.)

- 12.3** • If an ionic molecule results from the transfer of exactly one electron from one atom to the other, it should have a dipole moment $p = eR_0$, where R_0 is the bond

404 Chapter 12 • Molecules

length. Use the data in Table 12.1 (Section 12.2) to predict the dipole moments of KCl, LiF, NaBr, and NaCl in coulomb · meters. The observed values are, respectively, 3.42×10^{-29} , 2.11×10^{-29} , 3.04×10^{-29} , and 3.00×10^{-29} C · m; express these as percentages of your predicted values. (These percentages measure the extent to which the molecules concerned are ionic.)

- 12.4** • If a diatomic molecule is ionically bonded by the complete transfer of one electron, its dipole moment should be $p = eR_0$. Given the data in the table that follows, discuss the extent to which the molecules concerned are ionically bonded.

Molecule:	NaF	HF	CO
Bond length, R_0 (nm):	0.193	0.0917	0.113
Dipole moment, p (C · m):	2.72×10^{-29}	6.07×10^{-30}	3.66×10^{-31}

- 12.5** • A 1984 Ford Escort traveling at 50 mi/h experiences a drag force, because of rolling friction and air resistance, of about 360 N (roughly, 80 lb).* The energy released by 1 liter of gasoline is about 3.2×10^7 J and is used with about 18% efficiency. How far can the car travel on 1 liter of gasoline at 50 mi/h? Convert your answer to miles per gallon. (1 mi = 1.6 km, 1 gal = 3.8 liters.)

- 12.6** • In Example 12.2 (Section 12.2) we found that the energy released in the burning of 1 kg of carbon to give CO_2 is 9.1×10^7 J. In that calculation we assumed that the original carbon atoms were separate free atoms. [Specifically, the energy in Eq. (12.5) is for the case that the C atom is free and unattached.] If the C atoms are bound in a liquid or solid, some energy is needed to detach them, and the energy yield is smaller. Given that graphite (one of the solid forms of pure carbon) has a binding energy of 7.4 eV per atom, find the energy released when 1 kg of graphite burns in O_2 gas to form CO_2 . Compare your answer with the result in Example 12.2.

- 12.7** • Calculate the energy released if 1 kg of graphite (one of the solid forms of pure carbon) burns in oxygen gas to give carbon monoxide, in the reaction



Don't forget that energy is needed to dissociate the carbon atoms from the graphite and the oxygen atoms from the O_2 molecules. The binding energy of graphite is 7.4 eV per atom; the dissociation energy of the O_2 molecule is 5.1 eV, and that of CO is 11.1 eV.

- 12.8** • (a) The energy needed to dissociate the H_2O molecule into three separate atoms is 9.50 eV. Given that the dissociation energies of H_2 and O_2 are 4.48 and 5.12 eV, calculate the energy released in the reaction



*E. J. Horton and W. D. Compton, *Science*, vol. 225, p. 587 (1984).

- (b) The airship *Hindenburg* was held up by about 200 million liters of hydrogen gas. Assuming that this was at STP, estimate the energy released when the *Hindenburg* exploded. Give your answer in "tons of TNT." (One ton of TNT is 4.3×10^9 J, the energy released by 1 ton of the explosive TNT.)

- 12.9** • Gasoline is a variable mixture of various hydrocarbons, such as octane (C_8H_{18}). To be definite, consider pure octane and calculate the energy released by burning 1 kg of octane in the reaction



using the following dissociation energies:

Molecule:	C_8H_{18}	O_2	CO_2	H_2O
Dissociation energy (eV):	101.6	5.1	16.5	9.5

Note that these dissociation energies are the energies needed to separate each molecule completely into its individual constituent atoms. (Your answer here should be appreciably less than the value found in Example 12.2, which treated the oxidation of carbon atoms that were already separated.)

- 12.10** • Consider a point charge q outside a spherically symmetric, rigid charge distribution of zero total charge. (a) Use Gauss's law to prove that the distribution exerts no force on q . (b) Next, show that no assembly of external charges can exert a force on a spherically symmetric, rigid distribution of zero total charge.

- 12.11** • As a simple classical model of the covalent bond, suppose that an H_2 molecule is arranged symmetrically as shown in Fig. 12.34. Write down the total potential energy U of the four charges and, treating the protons as fixed, find the value of the electrons' separation s for which U is a minimum. Show that the minimum value is $U_{\min} \approx -4.2ke^2/R_0$. (Note that you have here treated the two protons as fixed at separation R_0 . Classically, they could not remain fixed in this way; it is one of the triumphs of quantum mechanics that it explains how the protons have a stable equilibrium separation, as we describe in Section 12.4. Note also that you should not take the number 4.2 in our answer too seriously — especially in estimating the dissociation energy. First, the electrons have appreciable kinetic energy, which we have not considered, second, even after the molecule has been dissociated into

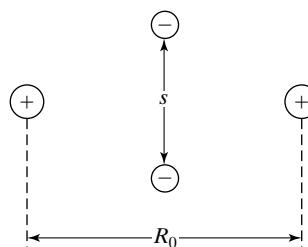


FIGURE 12.34

(Problem 12.11)

Problems for Chapter 12 **405**

two atoms, each atom still has an energy $-E_R$. Both of these effects lower the dissociation energy, and a more realistic estimate would be $B \sim ke^2/R_0$.)

SECTION 12.3 (The Ionic Bond)

- 12.12** • (a) Use the graphs in Fig. 10.11 to find the energy ΔE needed to transfer an electron between the following pairs of atoms: from Li to Br, from C to O, from Rb to F, from O to O. (b) Two of these pairs bond ionically and two do not; which are which?

- 12.13** • (a) Use the graphs in Fig. 10.11 to find the energy ΔE needed to transfer an electron between the following pairs of atoms: from Cs to I, from N to O, from N to N, from Na to Br. (b) Two of these pairs bond ionically and two do not; which are which?

- 12.14** • The ionization energy of K is 4.34 eV and the electron affinity of Cl is 3.62 eV. What is the critical radius R_c inside which the transfer of an electron from a K atom to a Cl is energetically favored?

- 12.15** • Use the ionization energies and electron affinities listed in Problem 12.18 to find the critical radius R_c inside which the transfer of an electron is energetically favored, for each of the ionic molecules KI, NaBr, LiF.

- 12.16** • The ionization energy of potassium is 4.34 eV, while the electron affinity of chlorine is 3.62 eV. Given that the equilibrium separation of KCl is 0.267 nm, estimate the dissociation energy of KCl using Eq. (12.10). Compare your answer with the observed value $B = 4.34$ eV.

- 12.17** • The ionization energy of Li is 5.4 eV, and the electron affinity of chlorine is 3.6 eV. Given that LiCl is ionically bonded with dissociation energy 4.8 eV, estimate the equilibrium separation R_0 of LiCl. Will your answer be an overestimate or underestimate?

- 12.18** • Use the following data to estimate the dissociation energies of the three ionic molecules KI, NaBr, and LiF. (IE denotes ionization energy and EA electron affinity.) Compare your answers with the observed values given in the last column. Your answers should all be overestimates; explain why.

IE (eV)	EA (eV)	R_0 (nm)	B (eV)
K: 4.34	I: 3.06	KI: 0.305	3.31
Na: 5.14	Br: 3.36	NaBr: 0.250	3.74
Li: 5.39	F: 3.40	LiF: 0.156	5.91

- 12.19** • Assuming that the following pairs of elements combine ionically, write down their chemical formulas: Ca and S, Ca and Cl, Li and O, Na and P. [Hint: Use the periodic table.]

- 12.20** • Assuming that the following pairs of elements combine ionically, write down their chemical formulas: Ba and S, Ca and I, Sr and Br, K and N. [Hint: Use the periodic table.]

- 12.21** • (a) Given that Ba and S bond ionically, what would you expect their valence to be? (b) Assuming that the bond is purely ionic, predict the electric dipole moment of BaS. (The measured bond length is $R_0 = 0.251$ nm.) (c) Express the observed dipole moment, 3.62×10^{-29} C·m, as a percentage of your answer in part (b).

SECTION 12.4 (The Covalent Bond)

- 12.22** • A crucial fact for the discussion of Section 12.4 is contained in Eq. (12.14), namely that if ψ_1 and ψ_2 satisfy the Schrödinger equation, then so does any linear combination $B\psi_1 + C\psi_2$. Starting with the standard three-dimensional Schrödinger equation (8.2), prove this result.

- 12.23** • Assuming that they bond covalently as described in Section 12.4, predict the chemical formulas of oxygen fluoride and nitrogen fluoride.

- 12.24** • Hydrogen can bond covalently with phosphorus, tellurium, and bromine. Predict the chemical formulas of the three resulting molecules.

- 12.25** • (a) Octane, C_8H_{18} , is called a *straight chain* hydrocarbon because its carbon atoms are arranged in a straight line. Draw a picture of the octane molecule showing all bonds, as in Fig. 12.13. (b) Do the same for the straight-chain propane, C_3H_8 . (In both cases all carbon atoms have valence 4.)

- 12.26** • Make a sketch similar to those in Fig. 12.13 of ethane C_2H_6 (in which both carbons have valence 4).

- 12.27** • Make a sketch similar to those in Fig. 12.13 of acetylene C_2H_2 (in which both carbons have valence 4). [Hint: Remember that more than one bond can join the two carbon atoms.]

- 12.28** • The observed electric dipole moments of CO, CIF, and NaF are 3.7×10^{-31} , 3.0×10^{-30} , and 2.7×10^{-29} C·m, respectively. The corresponding bond lengths are 0.11, 0.16, and 0.19 nm. Express the observed dipole moments as percentages of the values you would predict if the molecules were purely ionic. Which of the molecules are predominantly ionic, and which are predominantly covalent?

- 12.29** • (a) Assuming that the following pairs of elements combine covalently, predict the formulas of the resulting molecules.

	$p(\text{C} \cdot \text{m})$	R_0 (nm)
Cl and F	3.0×10^{-30}	0.16
Br and Cl	1.9×10^{-30}	0.21
I and Cl	4.1×10^{-30}	0.23

- (b) Use the observed dipole moments and bond lengths to confirm that these molecules are predominantly covalent.

406 Chapter 12 • Molecules

12.30•• (a) What should be the valence of Be in its ground state ($1s^2 2s^2$)? **(b)** As discussed in Section 10.6, the $2s$ and $2p$ levels are close together, and when another atom is nearby, it is often energetically favorable to promote one of the $2s$ electrons to a $2p$ state, so that a bond can form. What is the valence of Be in the configuration $1s^2 2s^1 2p^1$? **(c)** Predict the chemical formulas for the compounds of Be with fluorine, with oxygen, and with nitrogen.

12.31•• Consider the H_2^+ wave functions ψ_{\pm} discussed in Section 12.4. In that discussion we did not worry about normalization, but ψ_{\pm} should strictly have been defined as $\psi_+ = B(\psi_1 + \psi_2)$ and $\psi_- = C(\psi_1 - \psi_2)$, where B and C are normalization constants needed to ensure that $\int |\psi|^2 dV = 1$. **(a)** If ψ_1 and ψ_2 do not overlap (or, more precisely, if their overlap is negligible), show that $B = C = 1/\sqrt{2}$. (Assume that ψ_1 and ψ_2 are themselves normalized.) **(b)** If ψ_1 and ψ_2 overlap a little, argue that B is a little less than $1/\sqrt{2}$ and hence that at the midpoint between the two protons, $|\psi_+|^2$ is just a little less than $2|\psi_1|^2$. This proves our claim that ψ_+ concentrates the probability density between the two protons. **(c)** Argue similarly that C must be a little larger than $1/\sqrt{2}$.

SECTION 12.5 (Directional Properties of Covalent Bonds*)

12.32•• The water molecule is partially ionic, in that an electron is partially transferred from each hydrogen to the oxygen, as indicated in Fig. 12.35, where q denotes the magnitude of each of the two charges transferred. **(a)** Write down the electric dipole moment p of the H_2O molecule in terms of the charge q , the H—O bond length d , and the angle θ . (Note: The dipole moment of two charges q and $-q$ a distance d apart is a vector of magnitude qd pointing from the negative to the positive charge. Thus, H_2O has two dipoles \mathbf{p}_1 and \mathbf{p}_2 as shown, and the total moment is $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$.) **(b)** The measured values are $p = 6.46 \times 10^{-30} \text{ C} \cdot \text{m}$, $d = 0.0956 \text{ nm}$, and $\theta = 104.5^\circ$. Find the magnitude q of the charge transferred and express it as a fraction of the electron charge e .

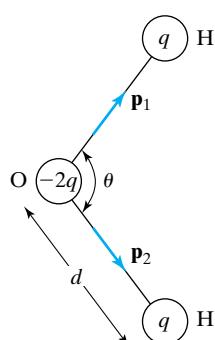


FIGURE 12.35

(Problem 12.32)

12.33•• We mentioned in Section 12.5 that the four bonds of a carbon atom (in a molecule like CH_4) point symmetrically to the four corners of a regular tetrahedron centered on the carbon nucleus. Prove that the angle between any two bonds is 109.5° .

SECTION 12.6 (Excited States of Molecules*)

12.34• The force constant for the HI molecule is $k = 2.0 \times 10^3 \text{ eV/nm}^2$. Calculate the vibrational energy of the lowest three vibrational levels of HI . (Treat the iodine atom as fixed.)

12.35• The separation of the lowest two vibrational levels in HBr is 0.33 eV . Find the force constant k of HBr . (Treat the bromine atom as fixed.)

12.36• Calculate the lowest four rotational energies of the HI molecule in eV, and sketch them in an energy level diagram. (The bond length of HI is 0.16 nm , and you can treat the iodine atom as stationary.)

12.37•• Figure 12.6 shows the energy $E(R)$ for the Na^+ and Cl^- ions that make up the $NaCl$ molecule. Near its minimum at R_0 , $E(R)$ can be approximated by a parabola: $E(R) \approx E(R_0) + \frac{1}{2}k(R - R_0)^2$. Use the information on the graph to make a rough estimate of the force constant k . [Hint: To the left of R_0 , the graph retains a reasonably parabolic shape as far as the point where it crosses the horizontal axis.]

12.38•• Consider an O_2 molecule rotating about its center of mass (as in Fig. 12.26). **(a)** Derive an expression for its moment of inertia I in terms of m , the mass of an O atom, and R_0 , the bond length. **(b)** Use Eq. (12.27) to find the lowest three rotational energy levels in eV. (The bond length of O_2 is 0.121 nm .)

12.39•• Consider two atoms of masses m and m' , bound a distance R_0 apart and rotating about their center of mass (as in Fig. 12.26). **(a)** Calculate their moment of inertia, I , and prove that it can be written as $I = \mu R_0^2$, where μ is the reduced mass

$$\mu = \frac{mm'}{m + m'} \quad (12.41)$$

This shows that one can treat the rotational motion of any diatomic molecule as if only one of the atoms were moving, provided that one takes its mass to be μ . **(b)** Prove that if $m \ll m'$, then $\mu \approx m$.

12.40•• Consider two classical atoms of mass m and m' , separated by a distance R , with their center of mass fixed at the origin. **(a)** Write down expressions for the atoms' distances, r and r' , from the origin, in terms of m , m' , and R . **(b)** Suppose that the two atoms vibrate in and out along the internuclear axis. Show that their total kinetic energy can be expressed as $K = \frac{1}{2}\mu(dR/dt)^2$, where μ is the reduced mass defined in Eq. (12.41) (Problem 12.39). This result proves for a classical molecule (what is true in quantum mechanics as well) that the vibrational motion of both atoms can be treated as if only one atom were moving, provided that we take its mass to be μ .

12.41 •• Consider the molecules HCl and DCl, where D denotes the deuterium, or heavy hydrogen, atom, $D = {}^2\text{H}$. **(a)** Explain why the force constant k and bond length R_0 should be about the same for DCl as for HCl. **(b)** The spacing of adjacent vibrational levels in HCl is $\hbar\omega_c \approx 0.37 \text{ eV}$. What is the corresponding spacing in DCl? (You may treat the Cl atom as fixed.)

12.42 •• The potential energy of two atoms that form a stable molecule has a minimum at a separation R_0 , as shown in Fig. 12.23. Consider two atoms with total kinetic energy K_0 approaching one another from far apart. **(a)** What will be their KE when they reach the equilibrium separation $R = R_0$? **(b)** At what point in the picture will the atoms come to rest? Will they remain there? **(c)** Assuming that the total energy remains entirely in the form KE + PE of the two atoms, can the two atoms come to rest at R_0 ? What value must R eventually approach? **(d)** We see that if the atoms are to form a stable molecule, some of their energy must be dissipated while they are near R_0 . Suggest a mechanism to do this.

12.43 ••• **(a)** Use the curve in Fig. 12.12 to get a rough estimate of the force constant k for the H_2^+ molecule. (See the hint for Problem 12.37.) **(b)** Use your value of k to estimate the zero-point energy of vibration in H_2^+ . (Since the two nuclei have equal mass, you will need to use the reduced mass as in Problem 12.40.)

12.44 ••• A simple analytic function that can be used to approximate the potential energy of the two atoms in a diatomic molecule is the Morse potential

$$U(R) = A[(e^{(R_0-R)/S} - 1)^2 - 1] \quad (12.42)$$

where A , R_0 , and S are positive constants, with $R_0 \gg S$. **(a)** Sketch this function for $0 \leq R < \infty$. **(b)** In terms of A , R_0 , and S , write down the bond length and binding energy of a molecule whose potential energy is given by (12.42). **(c)** Write down the first three terms of the Taylor series for $U(R)$ about R_0 (see Appendix B) and show that they give the parabolic approximation (12.25). **(d)** Show that the force constant k is $2A/S^2$.

SECTION 12.7 (Molecular Spectra*)

12.45 • The microwave spectrum of HCl consists of a series of equally spaced lines, $6.4 \times 10^{11} \text{ Hz}$ apart. Find the bond length of HCl (treating the Cl atom as fixed).

12.46 • The microwave spectrum of HBr consists of a series of equally spaced lines, $5.1 \times 10^{11} \text{ Hz}$ apart. Find the bond length of HBr (treating the Br atom as fixed).

12.47 • The vibrational–rotational spectrum (12.35) of a diatomic molecule is in the infrared region. At low resolution it is seen as a single band centered on $E_\gamma = \hbar\omega_c$. The molecule LiBr shows a bright band in the IR region, centered at $\lambda = 17.8 \mu\text{m}$. Use this to find the force constant k for LiBr (treating the Br atom as fixed).

12.48 •• A spectroscopist finds three lines in the far infrared spectrum of HI at 259, 195, and $156 \mu\text{m}$ (and no other lines in between). **(a)** Find the photon energies, and

identify the rotational transitions involved. **(b)** Hence, find the bond length of HI.

12.49 •• Repeat the calculation in Example 12.5 (Section 12.7) of the bond length of LiI, but do not make the approximation that the I atom is fixed. (Use the reduced mass introduced in Problem 12.39.) Compare your answer with that of Example 12.5, where we ignored the motion of the iodine atom.

12.50 •• A spectroscopist finds three lines in the microwave spectrum of CO at $\lambda = 2.59$, 1.29, and $0.86 \mu\text{m}$ (and no other lines in between). **(a)** Find the corresponding photon energies, and identify the transitions involved. **(b)** Using the reduced mass of Problem 12.39, find the bond length of CO.

12.51 •• The first four lines in the rotational spectrum of CaCl are shown schematically in Fig. 12.36. (The relative strengths of the four lines depend on the temperature and need not concern you here.) Use this information to find the bond length of CaCl. (Since the two masses are comparable, you will need to use the reduced mass introduced in Problem 12.39. Assume that the Cl is chlorine 35.)

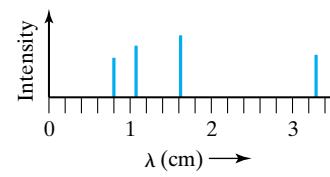


FIGURE 12.36

(Problem 12.51)

12.52 •• The vibrational–rotational spectrum of HI is sketched in Fig. 12.37. Use the information in that figure to find the force constant and bond length of HI.

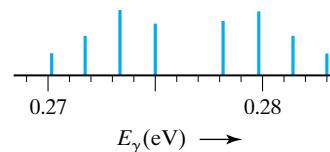
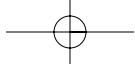


FIGURE 12.37

(Problem 12.52)

12.53 ••• **(a)** Consider an HCl molecule in which the Cl atom is ${}^{35}\text{Cl}$. Allowing for the motion of both atoms, find the wavelengths of the first two lines in the rotational spectrum. (Use the reduced mass introduced in Problem 12.39. Take $R_0 = 0.1270 \text{ nm}$, and keep four significant figures.) **(b)** Do the same for the case that the Cl atom is ${}^{37}\text{Cl}$. (Use the same value for R_0 .) **(c)** Make a sketch similar to Fig. 12.36 of these two lines in the emission spectrum of natural HCl, which contains 3 parts ${}^{35}\text{Cl}$ to 1 part ${}^{37}\text{Cl}$.



408 Chapter 12 • Molecules

COMPUTER PROBLEMS

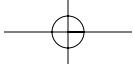
12.54 •• (Section 12.2) If you haven't already done Problem 12.11, do so now, and then make a plot of the potential energy $U(s)$ as a function of s . (For the purposes of the plot, you may as well choose units such that ke^2 and R_0 are equal to 1. Choose the ranges of U and s to show the interesting behavior.) Does your graph confirm the results of Problem 12.11?

12.55 •• (Section 12.6) The potential energy of the two atoms in HCl can be approximated by the Morse potential (12.42) (Problem 12.44) with

$$A = 4.57 \text{ eV}, \quad R_0 = 0.127 \text{ nm}, \quad S = 0.0549 \text{ nm}$$

(a) Plot the potential energy $U(R)$ for $0 \leq R \leq 0.3 \text{ nm}$. (Restrict the vertical axis to the range $-5 \leq U \leq 5 \text{ eV}$.) (b) What are the bond length and binding energy for HCl? (You may ignore the zero-point vibration.) (c) What is the force constant k for HCl? [See Problem 12.44(d).] (d) Plot the SHO well (12.25) that approximates $U(R)$ near R_0 . Superpose this plot on your previous one.

12.56 ••• (Section 12.4) As a simple model of the H_2^+ ion, consider the following one-dimensional system: An electron moves on the x axis, in the field of two identical nuclei at $x = \pm d$. Suppose that the wave functions for the electron bound to one nucleus, with the other nucleus removed, are $\psi_1 = Ae^{-(x+d)^2}$ and $\psi_2 = Ae^{-(x-d)^2}$. (a) Show that the normalization constant A is $(2/\pi)^{1/4}$, and plot the two wave functions for $d = 3$. The needed integrals are in Appendix B. (b) Now consider the bonding and antibonding wave functions $\psi_+ = B_+(\psi_1 + \psi_2)$ and $\psi_- = B_-(\psi_1 - \psi_2)$, and show that the normalization constants B_+ and B_- are equal to $1/\sqrt{2 \pm 2e^{-2d^2}}$, respectively. (c) Plot the wave functions ψ_+ and ψ_- for the case that $d = 3$ (with the nuclei reasonably far apart). Repeat for the probability densities $|\psi_+|^2$ and $|\psi_-|^2$. (d) Repeat all of part (c) but with the nuclei closer together at $d = 1$. Comment on your results, with particular reference to Figs. 12.9 and 12.10.



Chapter 13

Solids — Theory

- 13.1** Introduction
- 13.2** Bonding of Solids
- 13.3** Crystals and Noncrystals
- 13.4** Energy Levels of Electrons in a Solid; Bands
- 13.5** Conductors and Insulators — A Qualitative View
- 13.6** The Drude Model of Conductivity
- 13.7** Electron Collisions in Metals
- 13.8** The Fermi Speed
- 13.9** Degeneracy Pressure*
- 13.10** White Dwarfs, Neutron Stars, and Black Holes*
- 13.11** Classical and Quantum Gases*
- 13.12** Bose–Einstein Condensation*

Problems for Chapter 13

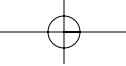
* These are advanced topics that can be omitted without serious loss of continuity.

13.1 Introduction

In Chapter 12 we saw how atoms can bond together to form molecules. In this chapter we study the aggregation of atoms on a larger scale, to form solids. There are many close parallels between solid-state and molecular physics. Several of the bonding mechanisms in solids are exactly the same as in molecules; for example, many solids are covalently or ionically bonded. There are, however, many differences between solid-state and molecular physics. For example, the bonding of metallic solids has no direct parallel in molecules. Also, certain solids, instead of being simple aggregates of atoms, are better viewed as being built in two stages, with atoms bound into molecules and then molecules bound into a solid. (A good example is ice, which is best understood as an aggregate of water molecules.) Another new feature of the solid state is the possibility that atoms (or molecules) can bond together into a symmetric, repetitive crystalline array; the study of the possible crystal structures is an important part of solid-state physics, with no exact counterpart in molecular physics.

Perhaps the most striking thing about solids is that they were the first example of *macroscopic* systems that were understood with the help of quantum mechanics.* Numerous commonplace properties of solids — specific heat, conductivity, transparency, magnetic susceptibility, compressibility, and many more — had defied classical explanation, but could be explained and calculated using quantum mechanics. It is this ability to explain — and hence control

*The applications of quantum theory to solids go back to the 1907 paper of Einstein, in which he used quantization to explain the specific heats of solids. See Section 15.9.



410 Chapter 13 • Solids — Theory

— the electronic and mechanical properties of solids that has led to a vast array of technological applications that have transformed civilization in the last century. Probably the most important technological advance in solid-state physics was the invention of the transistor, which led to miniaturized electronics, personal computers, and the digital information age we now live in. The transistor and several other applications of solid-state physics are examined in the next chapter. In this chapter we concentrate on the theory underlying these applications.

The theory of solids necessarily involves approximations. In principle, any property of any solid can be predicted using Schrödinger's equation, but in practice this strategy is fruitless. In the case of a macroscopic solid with some 10^{23} electrons and nuclei, Schrödinger's equation is exceedingly complex and impossible to solve exactly. (Even in the case of simple molecules with just a few atoms, Schrödinger's equation cannot be solved exactly.) The art of solid-state theory is to *model* the system under study by simply ignoring most of the real-world complexities and concentrate on just a few essential features. For instance, in computing the electronic properties of metal, one often assumes that the conduction electrons behave as free electrons in a box — thus the theorist completely ignores the interactions of the electrons with the nuclei and with each other. A theoretical model must be simple enough to allow calculations, but not so simple as to erase the properties the theorist wishes to explain. The check that the model is correct, that its approximations are valid, is always experiment.

We begin this chapter, in Sections 13.2–13.5, with a broad qualitative overview of the physics of solids. In Section 13.2 we describe the principal mechanisms by which atoms or molecules bond to one another to form solids. In Section 13.3 we describe some of the regular crystalline structures in which many solids can form, and we contrast these with the irregular, noncrystalline structure of the amorphous solids and the still more irregular composition of liquids and gases. We next take up what is perhaps the single most important achievement of solid-state physics, the theory of electrical conductivity. This theory depends on an understanding of the electron energy levels in a solid. In Section 13.4 we discuss these levels and describe how they fall into *bands* of allowed energies. Within these bands the energy levels are distributed almost continuously, but between them are *gaps* where there are no allowed levels. In Section 13.5 we show how the different possible arrangements of these bands explain why some solids are conductors, some are insulators, and some — the so-called semiconductors — are in between.

Following our introductory survey of the field, we take a closer look at the use of quantum mechanics in explaining properties of solids, with an emphasis on electronic properties. In Sections 13.6–13.8, we examine a relatively simple but very useful model of electrical conduction called the Drude model. We will see that at almost every step in our discussion of the Drude model, drastic approximations are made in order to keep the theory as simple as possible — and yet the theory agrees very well with experiment and explains many basic properties of metals and semiconductors.

We conclude the chapter with some further, somewhat exotic, uses of solid-state theory. Quantum mechanics explains why solids are so difficult to compress, compared with gases, and this effect, called degeneracy pressure, is explained in Section 13.9. In Section 13.10 we will see that degeneracy pressure has profound consequences in astrophysics. In the final two sections we describe a very strange state of matter that occurs only at very low temperatures, the Bose–Einstein condensate.

13.2 Bonding of Solids

Solids, like molecules, are held together by several different types of bonds. Just as with molecules, the distinctions between the various bonds are not always clear cut, and many solids are bonded by a combination of several mechanisms. Nevertheless, we can distinguish four main types of bonds — covalent, metallic, ionic, and dipole–dipole — and we describe each of these in turn.

Covalent Bonding

We saw in Chapter 12 that many molecules are held together by covalent bonds, each comprising two valence electrons shared by two of the atoms. In H_2 , for example, the two hydrogen atoms share their two electrons to form a single bond; in CH_4 (methane) each of the four valence electrons of the carbon joins with an electron from one of the four hydrogens to form a covalent bond. This same mechanism holds the atoms together in many solids. A beautiful example of such a covalent solid is diamond, which is one of the forms of solid carbon. In diamond each carbon atom is covalently bonded to four neighboring carbon atoms, as shown in Figure 13.1. The four bonds from any one carbon atom are directed symmetrically outward, and each atom in diamond is at the center of a regular tetrahedron formed by its four nearest neighbors. Both silicon and germanium — in the same group as carbon in the periodic table — are bonded covalently in a similar tetrahedral structure. Other examples of covalent solids are silicon carbide (widely used for its hardness and resistance to high temperatures) and zinc sulfide (used by Rutherford and other pioneers in nuclear physics as a scintillation detector).

An important characteristic of a solid is its strength. One simple measure of this strength is the **atomic cohesive energy**, defined as the energy needed, per atom, to separate the solid completely into individual atoms. In Table 13.1

TABLE 13.1

The cohesive energies of various solids. With two exceptions, the energies listed are atomic cohesive energies (energy per atom to separate the solid into individual atoms).* The exceptions are the molecular solids, ice and methane. Since these are properly considered assemblies of molecules, we list their molecular cohesive energies (the energy per molecule to separate the solid into individual molecules). The melting points shown in the last column can be seen to correlate well with the cohesive energies, as one would expect.

Bond	Solid	Cohesive energy	Melting point
Covalent	C (diamond)	7.37 eV/atom	>3770 K
	SiC	6.15 eV/atom	2870 K
	Ge	3.85 eV/atom	1231 K
Metallic	Fe	4.32 eV/atom	2082 K
	Cu	3.52 eV/atom	1631 K
	Pb	2.04 eV/atom	874 K
Ionic	LiF	4.16 eV/atom	1143 K
	NaCl	3.19 eV/atom	1074 K
	CsI	2.68 eV/atom	621 K
Dipole-dipole	H_2O	0.52 eV/molecule	273 K
	CH_4	0.10 eV/molecule	89 K
	Ar	0.08 eV/atom	84 K

*For solids containing two elements, like SiC and NaCl, the cohesive energy is often expressed in eV per atomic pair. With this convention, SiC would be shown as 12.3 eV per SiC pair, but this number is not directly comparable to the numbers listed for solids containing a single element, like diamond.

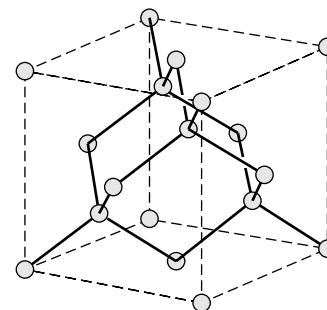
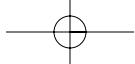


FIGURE 13.1

The structure of the diamond lattice. Each atom is at the center of a tetrahedron of four nearest-neighbor atoms.



412 Chapter 13 • Solids — Theory

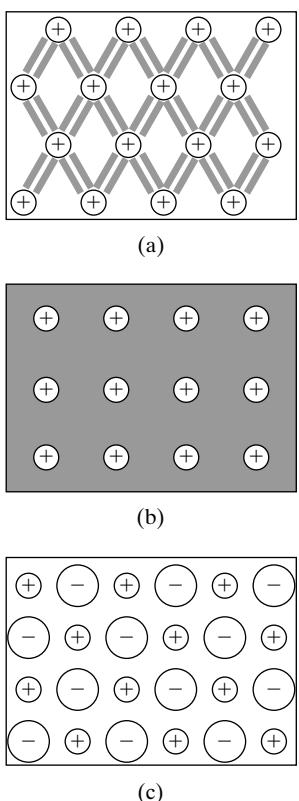
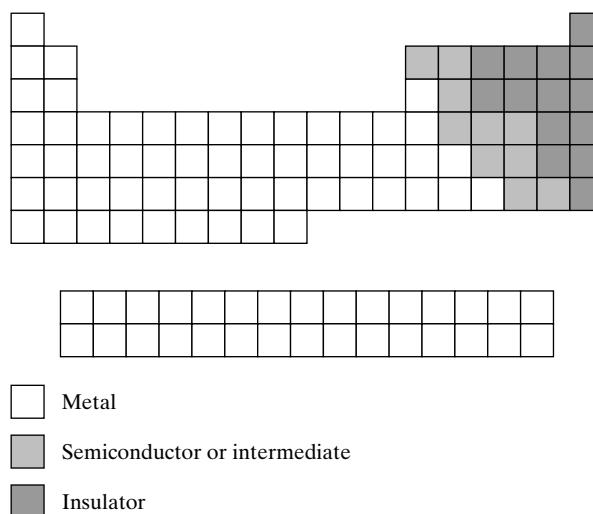


FIGURE 13.3

Bonding types in solids. **(a)** The covalent bond: pairs of electrons shared by two atoms. **(b)** The metallic bond: conduction electrons spread out nearly uniformly in the space between ion cores. **(c)** The ionic bond between positively and negatively charged closed-shell ions.

we have listed a number of solids with their cohesive energies. The first three entries are covalent solids, and we see what is generally true for covalent solids: Their cohesive energies vary from around 3 to about 7 eV per atom. This energy scale is not surprising, since chemical bonds involve the outer electrons in atoms, and a few eV is a typical energy for outer-shell electrons. As one would expect, those solids with high cohesive energies, like diamond and silicon carbide, are especially strong and have high melting points.

Since their electrons are usually well bound and cannot move easily, most covalent solids are poor conductors of electricity. In many covalent solids, the excitation energy of the electrons is higher than the energy of visible photons (2 or 3 eV). For this reason, many covalent solids (like diamond) cannot absorb visible photons and are therefore transparent to light.

Metallic Bonds

As seen in Figure 13.2, more than half of all elements are metals, and all solid metals are held together by the **metallic bond**. In some ways, this bond can be seen as a relative of the familiar covalent bond. In a covalent bond, electrons are shared between pairs of atoms. In the metallic bond, electrons are shared among *all* the atoms: One or two valence electrons are completely detached from each parent atom and can wander freely through the whole metal. Thus, a metal can be thought of as a lattice of positive atomic cores, immersed in a “sea” of mobile electrons, as in Figure 13.3(b). It is the attraction between the positive atomic cores and this negative sea of electrons that holds the metal together. As you can see in Table 13.1, the cohesive energy of typical metals is less than that of many covalent solids, but is, nevertheless, high enough to make most metals rather strong.

All metals are good conductors of electricity because the outermost one or two electrons per atom are free to move throughout the entire volume of the metal almost as if they were free electrons. For this reason, they are sometimes described as a “gas” of electrons. These free electrons are called **conduction electrons**, since they carry, or conduct, electrical current. Since the conduction electrons can also carry energy, metals are good thermal conductors as well. In addition, conduction electrons are the cause of the opaque but shiny appearance of metal surfaces. When a photon strikes a metal, it can easily

lose energy and momentum to the “free” electrons, and metals are therefore not transparent to light. In the language of classical electromagnetism, when visible light strikes the surface of a metal, the oscillating electric field of the light wave exerts an oscillating force on the electrons in the metal. Since the conduction electrons are free to move, they oscillate with large amplitude in response to the field, and their large oscillations create a new electromagnetic wave, the reflected wave, which produces the mirrorlike appearance of metals.

Ionic Bonds

We saw in Chapter 12 that certain pairs of elements can bond ionically to form molecules. For exactly the same reasons, these same pairs of elements can bond ionically to form solids. For example, in solid NaCl (common salt) each Na atom has lost one electron to a Cl atom; the resulting closed-shell ions arrange themselves so that each Na^+ ion is surrounded by several Cl^- ions and vice versa, as illustrated in Fig. 13.3(c).* The resulting solid is therefore quite tightly bound. As illustrated by the examples in Table 13.1, typical atomic cohesive energies (energy to separate the solid into neutral atoms) are 3 or 4 eV per atom for ionic solids.

Since the electrons of an ionic solid are all bound tightly in closed-shell ions, ionic solids are poor conductors of electricity and heat. For the same reason, visible photons cannot be absorbed easily and ionic solids are transparent to light.

Dipole-Dipole Bonds

The ionic bond results from the simple Coulomb force between oppositely charged ions. If the basic units (atoms or molecules) are electrically neutral, this force is not present. There is, nevertheless, a related, though weaker attraction that *can* act, even between neutral atoms or molecules. A neutral atom or molecule can have its charges distributed so that it has an electric *dipole moment*; that is, the preponderance of positive charge is centered at one point and that of negative, at another. Two neutral dipoles can exert an electrostatic force on one another, as we now describe.

Let us consider first two **permanent dipoles**, that is, two objects each of which is a dipole even when in complete isolation. (An example of a permanent dipole is the water molecule, as we saw in Chapter 12.) If we bring these dipoles together so that two like charges are closest, as in Figure 13.4(a), the repulsive forces are slightly greater than the attractive, and the net force is repulsive (See Problems 13.3 and 13.4). On the other hand, if the positive charge in one dipole is closest to the negative in the other, as in Fig. 13.4(b), the net force is attractive. In addition to these forces, each dipole exerts a torque on the other, and this torque tends to rotate them into the alignment of maximum attraction (c). Therefore, the net force between two permanent dipoles (that are free to rotate) is automatically attractive. Because it involves substantial cancellations between attractive and repulsive forces, this dipole-dipole attraction is much weaker than the Coulomb forces between the individual charges. Nevertheless, it is the attraction that holds many solids together.

Because of their basic spherical symmetry, no atoms are permanent dipoles. On the other hand, many molecules (the so-called *polar molecules*,

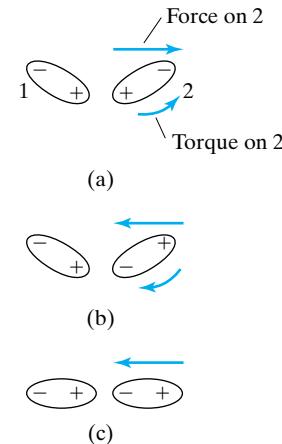
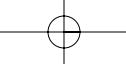


FIGURE 13.4

- (a) If two dipoles are oriented with like charges closest, they repel one another.
- (b) If unlike charges are closest, the dipoles attract.
- (c) The attraction is greatest when the dipoles are exactly aligned. In both (a) and (b) the dipoles also exert torques on one another, and these torques tend to rotate the dipoles into the attractive alignment of (c).

*Because each Na^+ “belongs” to several Cl^- ions and vice versa, you should not think of solid NaCl as made up of NaCl molecules; rather, it is a bonding of many Na ions with many Cl ions, forming one large crystalline array.



414 Chapter 13 • Solids — Theory

like H_2O) *are*, and these molecules can be bound into solids by the dipole-dipole attraction just described. Because the bond holding the molecules to one another is much weaker than the bond holding the atoms together inside each molecule, these solids are properly seen as composed of *molecules* and are sometimes called *molecular solids*. An example of a molecular solid is ice; we see from Table 13.1 that the energy needed to separate ice into H_2O molecules is about 0.5 eV per molecule. As expected, this is appreciably less than the cohesive energies of typical covalent, metallic, or ionic solids (usually several eV per atom). It is also much less than the 5 eV needed to remove any one atom from the H_2O molecule — which confirms that ice is properly regarded as an assembly of H_2O molecules.

The bond between two water molecules actually involves another effect as well. We saw in Section 12.5 that the oxygen atom in water appropriates at least part of the electron from each hydrogen, leaving nearly bare protons at the two ends of the molecule. When water molecules bond to form ice, these protons are shared with nearby oxygen atoms in somewhat the same way that electrons are shared in a covalent bond. This bonding by shared protons is called **hydrogen bonding** and plays an important role in biology: The DNA molecule comprises two intertwined helical strands of atoms. The atoms within each strand are bound together by covalent bonds, while the binding *between* the two strands is by hydrogen bonds, which are much weaker. This is what allows the two strands to unravel from one another without disturbing their internal structure in the process of replication.

Although all atoms and many molecules have zero permanent dipole moment, they can acquire an **induced moment** when put in the field of other charges. This is because the field pushes all positive charges one way and pulls all negatives the other. These induced dipoles are responsible for the very weak **van der Waals bond*** between some atoms and molecules, as we describe shortly. Because this bond is so weak, it is important only in solids where all other types of bond are absent. For example, the noble gas atoms are subject to none of the bonding mechanisms described earlier, but can form solids because of the van der Waals attraction. Because the bond is so feeble, the noble gases solidify only at rather low temperatures. For instance, as shown in Table 13.1, solid argon has a cohesive energy of only 0.08 eV/atom, and its melting point is 84 K. The van der Waals force is also responsible for the solids of certain molecules that have no permanent dipole moment, such as O_2 , N_2 , and CH_4 (methane).

The origin of the van der Waals force between two atoms (or molecules) that have no permanent dipole moments can be explained by the following classical argument: A classical atom consists of Z electrons orbiting around a fixed nucleus. The statement that it has no permanent dipole moment means (classically) that *when averaged over time* the dipole moment is zero. Nevertheless, as the electrons move around, they can produce an instantaneous nonzero dipole moment. Now consider two such atoms close together. Suppose that at a certain instant, atom 1 has a nonzero dipole moment. This produces an electric field, which induces a dipole moment in atom 2. This induced dipole is always aligned such that it is *attracted* by atom 1. (This is illustrated, for one possible orientation of the atoms, in Figure 13.5. For other orientations,

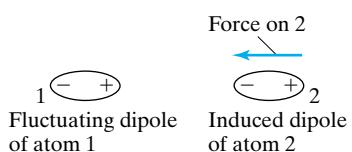


FIGURE 13.5

The fluctuating dipole moment of atom 1 induces a dipole moment in atom 2 (and vice versa). The direction of the induced moment is such that the two dipoles attract one another.

*The terminology here is a little confused. A few authors use “van der Waals” to describe *both* the attraction between two permanent dipoles *and* the weaker attraction involving induced dipoles. We follow the majority and reserve “van der Waals” to describe only the weaker induced dipole attraction.

Section 13.3 • Crystals and Noncrystals **415**

see Problem 13.5.) Therefore, as the two dipoles fluctuate, each induces in the other a dipole that it attracts. This is the origin of the van der Waals attraction between any two atoms or molecules.

13.3 Crystals and Noncrystals

All the mechanisms that bond atoms (or molecules) together into solids have the same general behavior: At large separations, r , the force between two bonding atoms is attractive; at a certain separation, $r = R$, it is zero; and for r less than R , it is repulsive. This behavior is illustrated in Fig. 13.6, which shows the potential energy of two atoms as a function of their separation r .

When just two atoms interact as in Fig. 13.6, they have minimum possible energy, and hence form a stable bound state at separation R . Three such atoms would be in equilibrium at the corners of an equilateral triangle, with the distance between each pair equal to R . Similarly, four would be in equilibrium at the corners of a regular tetrahedron. However, with five or more atoms, it is impossible to arrange them so that the distances between all pairs are the same. Instead, the atoms must arrange themselves so that any one atom is closest to a few others and farther away from all the rest. It is one of the tasks of solid-state physics to predict, or at least explain, the arrangement of the many atoms in a solid that minimizes their total energy and hence makes the solid stable.

As one might expect, the stable configuration of many solids has the atoms arranged in a regular pattern, in which the grouping of atoms is repeated over and over again. This regular pattern is called a **lattice**, and the resulting solid is called a **crystal**. Experimentally, the structure of crystals is found using diffraction of X-rays, neutrons, and electrons, and by the use of various kinds of microscopes (Figure 13.7). The mathematical study of the possible crystal structures is an important part of theoretical solid-state physics. Here we will just describe a few such structures.

The simplest crystal structure is the **cubic lattice**, in which the atoms are arranged at the corners of many adjacent cubes, as shown in Fig. 13.8. In a perfect cubic lattice all these cubes are identical, and each is placed in the same relation to its neighbors. Therefore, all the information about the structure of the lattice is contained in the single cube, or **unit cell**, shown in Figure 13.8(c).

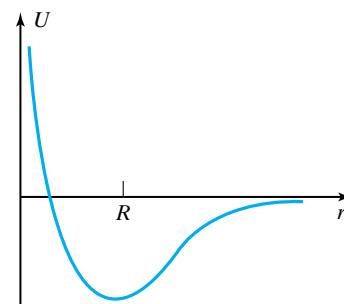


FIGURE 13.6

The potential energy of two atoms (or molecules) that can bond to form a solid is shown as a function of their separation r . Their equilibrium separation is labeled R .

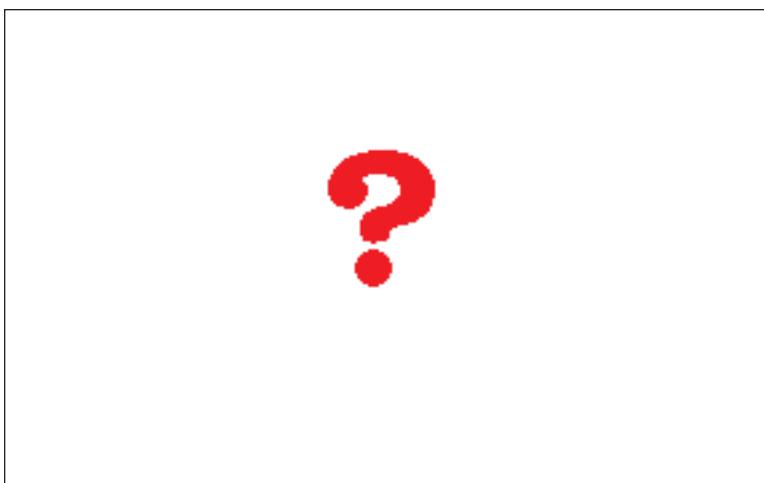
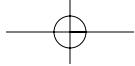


FIGURE 13.7

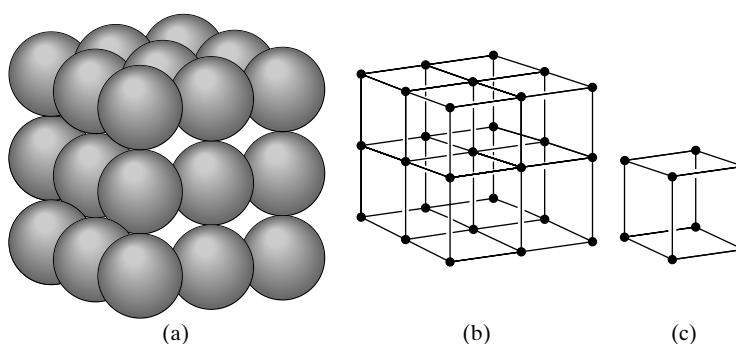
An image of the surface of graphite made with a scanning tunneling microscope, or STM. Each peak represents a single atom, and the regular crystalline pattern, with the atoms arranged in hexagons in this case, is clearly visible. The STM measures the current of electrons that “tunnel” between the sample and a sharp needle that is scanned across the sample’s surface. Since the current varies strongly with the distance between the surface and the needle, this produces a map of the surface, with a resolution that is much less than an atomic diameter. The scale markings are in angstroms ($1 \text{ \AA} = 0.1 \text{ nm}$).



416 Chapter 13 • Solids — Theory

FIGURE 13.8

(a) A simple cubic lattice made from 27 spherical atoms. (b) The structure of the lattice is much clearer if one shows just the centers of the atoms, using black dots, and connects the dots with lines. Although the actual appearance of the atoms in a crystal is more like (a) than (b), we will use the dot-and-line arrangement of (b) because of its greater clarity. (c) The unit cell of the same lattice is a single cube. The whole crystal can be constructed from many duplicates of this unit cell. The lines do not necessarily represent bonds, but help us visualize the geometry of the structure.



A slightly more complicated crystal lattice is the **face-centered cubic** (or **FCC**) lattice, with atoms on each face of a cube as well as at the corners, as in Figure 13.9(a). This arrangement actually gives a slightly higher density of atoms (more atoms per unit volume) than the simple cubic and is therefore favored in many solids. For example, common salt (NaCl) is an FCC crystal. As shown in Figure 13.9(b), the Cl^- ions are arranged on one FCC lattice. The Na^+ ions lie on a second FCC lattice, which is offset by half the height of the unit cube. In this way, each Na^+ is nearest to six Cl^- ions and vice versa.

Example 13.1

Given that the nearest-neighbor separation in NaCl is $r_0 = 0.28 \text{ nm}$, estimate the atomic cohesive energy of NaCl in eV per atom.

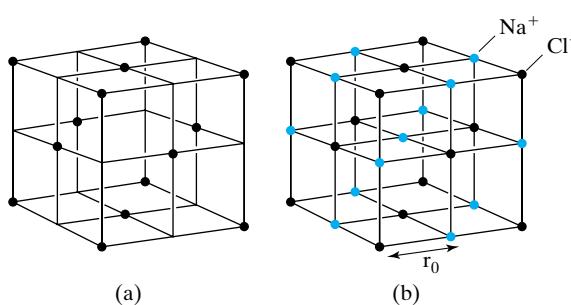
Let us consider first the potential energy of any one ion, for instance, the Na^+ ion at the center of the cube in Figure 13.9(b). This ion has six nearest-neighbor Cl^- ions, which give it a negative potential energy of $-6ke^2/r_0$. However, the Na^+ ion also has 12 next-nearest-neighbor Na^+ ions at a distance of $r_0\sqrt{2}$. (These are all of the remaining 12 Na^+ ions in the picture.) These contribute a positive potential energy of $+12ke^2/(r_0\sqrt{2})$. Next, there are eight Cl^- ions (at the eight corners of the cube in the picture) at a distance of $r_0\sqrt{3}$, and these contribute a negative potential energy of $-8ke^2/(r_0\sqrt{3})$. Continuing in this way, we would find a total potential energy:

$$\text{total PE of any one ion} = -\frac{ke^2}{r_0} \left(6 - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \dots \right) = -\alpha \frac{ke^2}{r_0} \quad (13.1)$$

The sum of the infinite series, α , is called the **Madelung constant**. Since the series is difficult to evaluate, we will simply state the result, that its sum is

FIGURE 13.9

(a) A unit cell of the face-centered cubic (or FCC) lattice. There is an atom at each corner of the cube and another at the center of each face. (b) The NaCl structure. The Cl^- ions form one FCC lattice, and the Na^+ ions another, offset by half the height of the unit cube. Each Na^+ is nearest to six Cl^- ions and vice versa. (This is most easily seen for the Na^+ at the center of the whole cube.) The distance r_0 is the nearest-neighbor distance between any ion and its six neighboring ions of opposite sign.



Section 13.3 • Crystals and Noncrystals **417**

$\alpha = 1.75$ (to three significant figures). Therefore, the potential energy of any one ion in solid NaCl is

$$\begin{aligned} \text{total PE of any one ion} &= -\alpha \frac{ke^2}{r_0} = -1.75 \times \frac{1.44 \text{ eV} \cdot \text{nm}}{0.28 \text{ nm}} \\ &= -9.0 \text{ eV} \end{aligned} \quad (13.2)$$

In this calculation we have ignored the short-range repulsive forces due to the overlap of the electron distributions when the ions get very close. The repulsive contribution to the energy is actually rather small (for exactly the same reason as it was in ionic molecules; see Section 12.3). Therefore, our estimate (13.2) is just a little more negative than the observed value

$$\text{total PE of any one ion} = -7.9 \text{ eV} \quad (13.3)$$

For the remainder of this calculation, we will use this observed value.

To find the energy of the whole NaCl crystal, we cannot simply multiply the answer (13.3) by the total number (N) of ions, since this would count each interaction twice over.* Instead, we must multiply by $N/2$, to give a total potential energy of $(-7.9 \text{ eV})N/2$. Thus the energy to pull the solid NaCl apart into separated Na^+ and Cl^- ions is $(7.9 \text{ eV})N/2$. Since $N/2$ is the number of $\text{Na}^+ - \text{Cl}^-$ pairs, we can say that

$$\text{energy to separate NaCl into } \text{Na}^+ \text{ and } \text{Cl}^- \text{ ions} = 7.9 \text{ eV per } \text{Na}^+ - \text{Cl}^- \text{ pair}$$

If we wish to have separate *neutral atoms*, we must next remove an electron from each Cl^- ion (which requires 3.6 eV), and then attach it to the Na^+ ion (which gives us back 5.1 eV). Thus, the total work done is

$$\begin{aligned} \text{energy to separate NaCl into Na and Cl atoms} &= (7.9 + 3.6 - 5.1) \text{ eV per pair} \\ &= 6.4 \text{ eV per Na-Cl pair} \end{aligned}$$

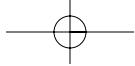
Dividing by 2, we find finally that

$$\text{atomic cohesive energy of NaCl} = 3.2 \text{ eV/atom}$$

Many other solids have the FCC lattice, for example, diamond[†] and several metals, including copper, silver, and gold. Two other important structures are the body-centered cubic (BCC) and the hexagonal close-packed (HCP), both sketched in Figure 13.10. Many metals, such as sodium, potassium, and iron, are BCC; examples of HCP are the solid forms of hydrogen and helium, and several metals, such as magnesium and zinc.

*For each pair of ions, i and j , there is a single potential energy of interaction that went into (13.3). However, if we simply multiply (13.3) by N , we will have counted the potential energy of i due to j and of j due to i . In other words, we will have counted every interaction twice.

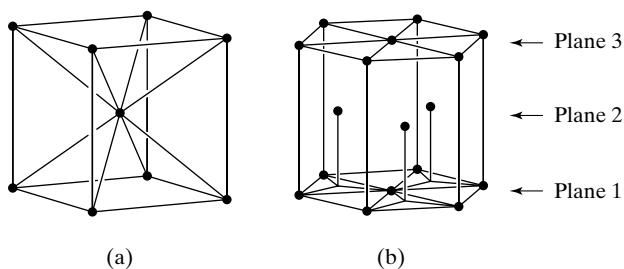
[†]This claim does not contradict our earlier statement that diamond has a tetrahedral structure. The carbon atoms are placed on two FCC lattices, one offset from the other. Each atom on either lattice is covalently bonded to its four nearest neighbor atoms, all of which are on the other lattice at the corners of a regular tetrahedron.



418 Chapter 13 • Solids — Theory

FIGURE 13.10

(a) The body-centered cubic (BCC) lattice. The unit cell has one atom at each corner of a cube and one at the center. (b) The hexagonal close-packed structure, which consists of equally spaced planes, each of which is closely packed in a hexagonal pattern. Planes 1, 3, 5, ..., are aligned vertically, as are planes 2, 4, 6, ..., but the odd planes are all displaced sideways from the even planes. Plane 2 has exactly the same arrangement as 1 and 3, but only three of its atoms fall inside the unit cell shown.



Macroscopic Appearance of Crystals

When a large enough number of atoms form into a single crystalline lattice, the resulting solid is a macroscopic crystal. As the crystal forms, its macroscopic surfaces tend to coincide with the planes of its microscopic lattice; if a grown crystal is struck, it tends to cleave along these same atomic planes. Either way, the crystal acquires smooth flat surfaces that reflect light like a mirror. If the material is transparent, these same surfaces can cause strong internal reflections, which add to the glittering appearance that we associate with crystals. There are many examples of such macroscopic crystals: ordinary table salt, gemstones, the fluorite crystal in Figure 13.11, and many more.

Many solids, including most metals, have a microscopic crystalline structure but do not, nevertheless, *appear* crystalline. This is because instead of forming as a single large crystal, these solids form in a jumbled array of many tiny microcrystals, called **grains**, as illustrated in the micrograph of stainless steel in Figure 13.12. Since a typical microcrystal, or grain, is about $1 \mu\text{m}$ across, the crystalline nature of such solids is completely invisible to the naked eye. Nevertheless, each grain contains some 10^{10} atoms, and, from a microscopic point of view, the dominant feature of the material is its crystalline structure.

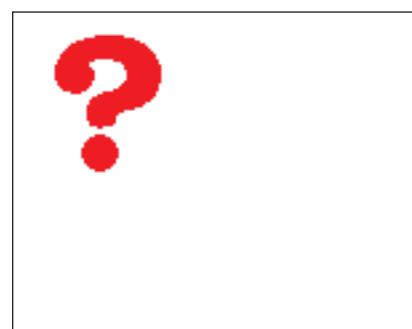


FIGURE 13.11

A crystal of fluorite (CaF_2). Note the well-defined faces and edges.

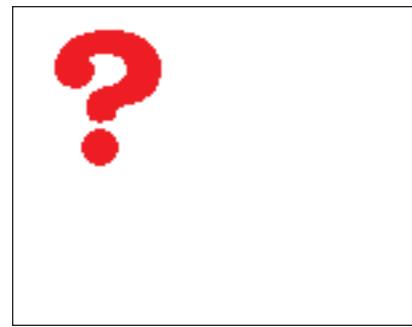


FIGURE 13.12

When magnified 2500 times, stainless steel is seen to be a disordered assembly of many microcrystals, called **grains**, each a few μm across.

Example 13.2

An important task for solid-state physics is to predict the strengths of materials. Given the estimate in Problem 13.22 for the force needed to separate two bonded atoms, make a rough estimate of the tensile strength of a simple cubic material with lattice spacing $r_0 \approx 0.3 \text{ nm}$.

The tensile strength of a material is the force per unit area needed to rupture it. From Problem 13.22 we find that the force needed to separate a single pair of bonded atoms is of order $F \approx 3 \times 10^{-9} \text{ newton}$. The density of atoms in any one crystal plane is $1/r_0^2$. Therefore if we ignore all but nearest-neighbor interactions, the force to tear apart a unit area of two adjacent planes is F/r_0^2 ; that is

$$\text{tensile strength} = \frac{F}{r_0^2} \approx \frac{3 \times 10^{-9} \text{ N}}{(3 \times 10^{-10} \text{ m})^2} \approx 3 \times 10^{10} \text{ N/m}^2 \quad (13.4)$$

This calculation assumed a cubic lattice and included only nearest-neighbor interactions. Furthermore, although the force that we used is realistic for strong bonds (covalent, ionic, or metallic), it would be a gross overestimate for weaker bonds such as the van der Waals bond. Nevertheless, the answer (13.4) should — and does — give the right order of magnitude for any single crystal of strongly bonded atoms. Single-crystal whiskers of iron (as in Figure 13.13) and of quartz have tensile strengths of this order. Macroscopic samples of many crystalline materials have tensile strengths much smaller than (13.4) (that for cast iron is some 100 times smaller), but this is because such macroscopic samples are not perfect crystals. It is the many imperfections in these materials that break easily and explain their much lower tensile strengths.

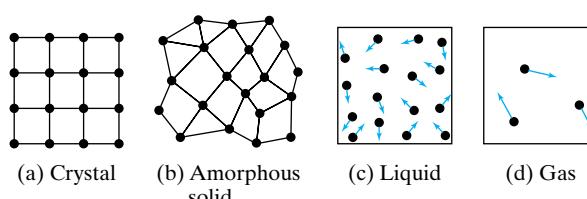
**FIGURE 13.13**

These nearly perfect crystalline whiskers of iron (shown magnified about six times) have tensile strengths of order 10^{10} N/m^2 in agreement with (13.4).

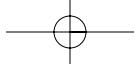
Amorphous Solids

Not all solids have a crystalline structure. Instead, the atoms of many solids, such as glass, wax, and rubber, arrange themselves in an irregular pattern. Such solids are described as *amorphous*, from a Greek word meaning “without form.” In fact, many materials can solidify in either a crystalline or an amorphous form, depending on their preparation. For example, a material may form in a crystalline state if cooled slowly from a liquid, a process called **annealing**. (Slow cooling gives the atoms time to arrange themselves in the lowest-energy, crystalline, state.) On the other hand, the same material may form in an amorphous state if **quenched**, that is, cooled very rapidly.

The difference between crystalline and amorphous solids is shown schematically in Figure 13.14. Part (a) illustrates a perfect crystal, with its atoms in a rigid and regular lattice. Part (b) shows an amorphous solid, with its atoms bonded in a more-or-less rigid structure, but not in a perfectly regular

**FIGURE 13.14**

Schematic representation of atoms (or molecules) of a given type in (a) a crystal, (b) an amorphous solid, (c) a liquid, and (d) a gas. The atoms of the solids (a) and (b) are bound in a rigid lattice, as indicated by the lines, which represent the permanent interatomic bonds. The atoms of the liquid and gas, (c) and (d), are free to move around, as indicated by the arrows; for this reason, a liquid or gas requires a container to confine it. (The atoms of a solid can also move, but only in small-amplitude oscillations about their equilibrium positions.) Although the atoms in the liquid (c) are not rigidly bonded, they remain within the range of the interatomic forces.



420 Chapter 13 • Solids — Theory

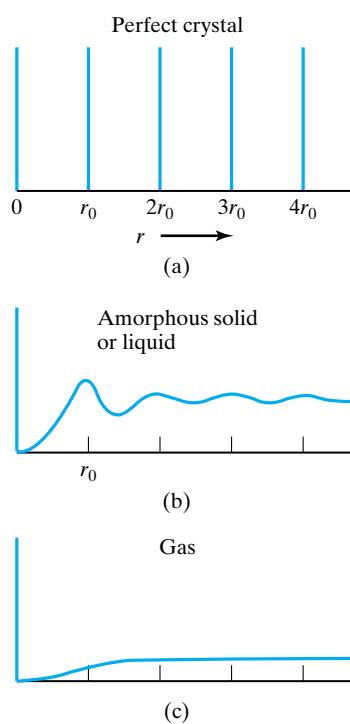


FIGURE 13.15

The probability densities for (a) a perfect crystal, (b) an amorphous solid or a liquid, and (c) a gas, all in one dimension. In each case the origin is placed at the known location of one atom. The probability of finding an atom is plotted up, and the distance away from the original atom is plotted to the right.

pattern. For comparison, parts (c) and (d) illustrate a liquid and a gas. When one heats a solid sufficiently, it first melts as in part (c): Although its atoms remain within the range of interatomic forces, they are no longer held in a rigid pattern and are, instead, free to move around. When heated still further, the liquid becomes a gas as shown in part (d): The atoms separate much further, they move more rapidly, and they interact only during their brief collisions.

We can characterize the differences between crystals, amorphous solids, liquids, and gases in terms of their **long-range order**. A perfect crystal has perfect long-range order: If we know its crystal structure and the location of one atom, we can predict with certainty the location of all other atoms. This is illustrated in Figure 13.15(a), which shows the probability density (the probability of finding an atom) plotted along one crystal axis for an ideal crystal. If we know that there is an atom at the origin, we can predict with certainty that there are atoms at r_0 , $2r_0$, $3r_0$, \dots , for as far as the solid extends.*

The distribution function for an amorphous solid is like Figure 13.15(b). If we know that there is an atom at the origin, the probability of finding another atom very close to $r = 0$ is small, since two atoms repel one another strongly at short range. The probability increases to a peak at about the same r_0 at which the next atom of a crystal would be located. It then drops again because the repulsion between overlapping atoms prevents another atom from locating too close to the atom of the first peak. Because of the amorphous solid's irregular structure, we can predict the atoms' positions with less and less certainty as we move on, and the density function rapidly approaches a constant, reflecting our complete ignorance of the positions of the more distant atoms. We say that the amorphous solid shows short-range order, but no long-range order.

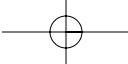
The probability density for a liquid is almost exactly like that for an amorphous solid, shown in Figure 13.15(b). In fact, an amorphous solid can be regarded as liquid, "frozen in time." The probability density for a gas is shown in part (c). It is much lower than those for solids or liquids, since the density is much less; it shows even less structure than that for the amorphous solid because a gas has even less order.

13.4 Energy Levels of Electrons in a Solid; Bands

We have stated that the reason metals conduct electricity is that some of their conduction electrons can move essentially freely within the boundaries of the metal, but this does not explain *why* the electrons in a metal can move so freely nor why the electrons in an insulator cannot. To explain this difference, we must examine the energy levels for the electrons in a solid. Fortunately, we can do this quite easily, leaning on our discussion of the energy levels of electrons in molecules (Section 12.4). As one might expect, once we understand the energy levels of the electrons, we can explain many properties of solids in addition to their conductivity.

In Section 12.4 we described the energy levels of an electron belonging to either of two atoms as the atoms come together to form a molecule. While the atoms are far apart, we can focus on any one atomic orbital ($1s$, $2s$, \dots).

*Strictly speaking, Fig. 13.15(a) represents a classical crystal at absolute zero. In quantum mechanics the uncertainty principle means that we cannot know the locations with complete certainty, and at nonzero temperatures thermal motions introduce a further uncertainty. For both reasons the ideal spikes of Fig. 13.15(a) should be a little spread out.

Section 13.4 • Energy Levels of Electrons in a Solid; Bands **421**

Since the electron can occupy this orbital in either atom, this gives two degenerate orbitals (that is, two possible wave functions with the same energy). We saw in Section 12.4 that as the two atoms come closer together and begin to overlap, this degeneracy splits and we have, instead, two orbitals of different energy, as indicated in Figure 13.16(a).

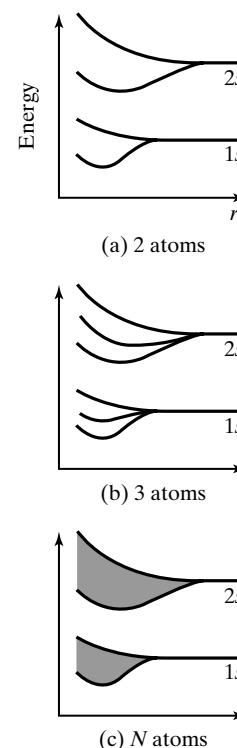
If we start instead with *three* atoms far apart, then for each atomic orbital, there are *three* independent, degenerate wave functions for an electron (corresponding to attaching the electron to any of the three atoms). As we move the atoms together, this threefold degeneracy is split, and we get three distinct possible energies, as indicated in Figure 13.16(b). More generally, if we consider *N* atoms far apart, then for each atomic orbital, there are *N* degenerate wave functions; and as we move the atoms together, these fan out into *N* more-or-less evenly spaced levels. We have shown this in Figure 13.16(c) where we have indicated the many closely spaced levels by shading.

We can imagine putting together a whole solid in this way, starting with *N* well-separated atoms. While they are still far apart, we can arrange the atoms in the same geometrical pattern as the final solid, and we can then imagine the whole structure slowly contracting to its equilibrium size. As the nearest-neighbor separation *r* decreases, the energy levels of the electrons will vary, their general behavior being that shown in Figure 13.16(c), which we have redrawn in more detail as Figure 13.17. For each atomic orbital, there are *N* independent wave functions, all of which have the same energy when *r* is large. As we reduce *r*, this degenerate level fans out into *N* distinct levels. We refer to these *N* closely spaced levels, which all evolved from a single atomic level, as a **band**. The width and spacing of these bands varies with *r*, but the bands of the actual solid are found by setting *r* equal to its equilibrium value *r*₀, as shown on the left of Figure 13.17.

Figure 13.17 introduces some of the terminology used in connection with the energy levels of electrons in solids. The band of levels that evolved from the 1s atomic levels is called the 1s band, and similarly with 2s, 2p, and so on. The range of energies in a single band — the *width* of the band — is typically a few eV. If there are *N* levels within the band,* the average spacing δE between adjacent levels is

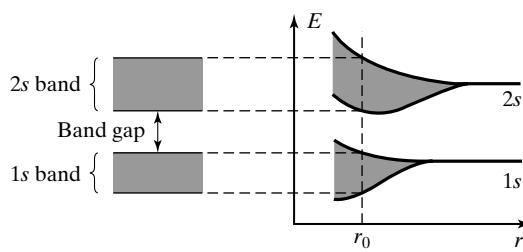
$$\delta E \approx \frac{\text{a few eV}}{N}$$

In practice, *N* is very large; even in a microcrystal 1 μm across, there are some 10^{10} atoms (Problem 13.25). Thus, the level spacing within a band is very small; in a microcrystal with $N \approx 10^{10}$, δE is of order 10^{-10} eV, and in a larger crystal it is even smaller. Therefore, for almost all practical purposes, we can

**FIGURE 13.16**

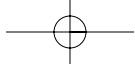
Energy levels of an electron in a system of several atoms, shown as functions of the atomic separation *r*.

- (a) For two atoms, each atomic level splits into two levels as *r* is reduced.
- (b) With three atoms, each level splits into three levels.
- (c) With many atoms (*N*, say), each atomic level fans out into *N* levels, which are so closely spaced that we can regard them as a continuum.

**FIGURE 13.17**

As we assemble a solid from *N* well-separated atoms, the original atomic energy levels fan out into *bands* of closely spaced levels. The bands of the actual solid are shown on the left and are found by taking the nearest-neighbor distance *r* equal to its equilibrium value *r*₀.

*As we will see shortly, there are generally more than *N* states in a band, but this only strengthens our conclusion that the average spacing is very small.



422 Chapter 13 • Solids — Theory

think of the electron energy levels in a solid as *continuously distributed within each band*. This is what the shading of the bands in Figure 13.17 was intended to suggest.

It can happen that the highest level in one band is above the lowest level in the next-higher band. In this case we say that the two bands overlap. On the other hand it often happens (as in Fig. 13.17) that two neighboring bands do *not* overlap, and there is instead a **band gap** — a range of energies that are not allowed. We will see that the occurrence of band gaps and their widths are two of the main factors that determine the conductive properties of a solid.

It is important to know whether all the states in a given band of a solid are occupied (just as it was important in our discussion of the periodic table to know whether all the states of a given level in an atom were occupied). To this end, we need to know how many states there are in a given band. For *s* bands (the 1s and 2s bands of Figure 13.17, for example) there are N independent wave functions and two possible spin orientations, and hence $2N$ states in all. If we consider a *p* band (a band that evolved from an $l = 1$ atomic level), then when the N atoms were well separated there were $3N$ independent wave functions (three different values of l_z for each atom) and two spin orientations. Therefore, there are altogether $6N$ states in a *p* band. Since we will be concerned only with *s* and *p* bands, we will not give a general formula here (but see Problem 13.28).

13.5 Conductors and Insulators — A Qualitative View

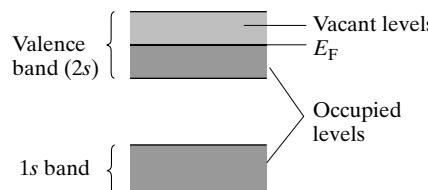
We are now ready to explain the difference between conductors and insulators. We illustrate our discussion with three examples, starting with lithium, which we know to be a metal and hence a conductor.

Lithium: A Conductor

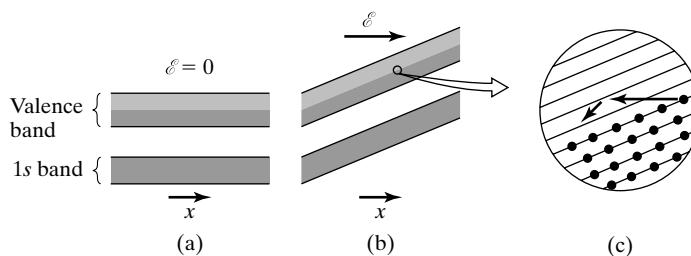
The lithium atom has the configuration $1s^22s^1$, with its 1s level full and its 2s valence level half-full. As we assemble N lithium atoms into solid Li, the 1s and 2s levels fan out into the 1s and 2s *bands*, which are shown in Figure 13.18. The N lithium atoms contribute altogether $3N$ electrons, while each of the two bands can hold $2N$ electrons. Thus the 1s band is completely filled, and the 2s *valence band* is half-filled. We show this occupancy in Figure 13.18, where dark shading represents occupied and light shading unoccupied levels. The energy shown as E_F is the **Fermi energy**, which is defined as the energy of the highest occupied level.*

FIGURE 13.18

The bands of allowed electron energies in solid lithium. All levels in the 1s band are *fully occupied* (dark shading); the 2s valence band is only *half full*, with empty levels shown lightly shaded.



*Strictly speaking, this is correct only at absolute zero. At higher temperatures, thermal motions can excite a few electrons to levels somewhat above E_F as described in Section 13.11.

Section 13.5 • Conductors and Insulators — A Qualitative View **423****FIGURE 13.19**

Schematic representation of the electron energy levels along a solid Li wire. (a) In the absence of an electric field, the levels are the same at all points along the wire. (b) An electric field in the x -direction raises the potential energy (and hence the total energy) on the right and tilts the whole diagram as shown. (c) A magnified portion of the valence band of (b). In response to the electric field, those electrons in the highest occupied levels can now move into unoccupied levels on their left, and a current flows. The electrons eventually collide with atoms of the lattice and lose energy, but the process of accelerations followed by collisions repeats itself for as long as there is an electric field. (Although we have drawn the levels of part (c) as discrete levels, for all practical purposes, they are a continuum.)

We are now going to argue that the electrons in the partially full valence band can move around in response to an applied electric field — they are conduction electrons, so the partially full valence band is also called the *conduction band* — whereas the electrons in the full 1s band cannot move. To see this, let us consider a lithium wire oriented along the x axis. In the absence of an electric field, the energy levels are the same at all points along the wire, as indicated in Figure 13.19(a).

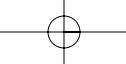
Suppose now that we connect a battery to the ends of the wire to maintain an electric field to the right. This field raises the potential energy of the electrons on the right, and the bands now look like Figure 13.19(b) or, in a magnified version, Figure 13.19(c). For any electron at the top of the occupied levels in the conduction band, there are now *unoccupied* levels of the same energy (and lower) just to the left. Therefore, each of these electrons begins to move to the left in response to the electric field, as indicated (for one electron) by the horizontal arrow in (c). This motion of electrons to the left is, of course, the expected electric current. In due course, the electrons moving to the left collide with defects and other “imperfections” of the lattice and lose some kinetic energy, as indicated by the downward arrow in Figure 13.19(c). These collisions, which are discussed in Section 13.7, are the main cause of electrical resistance in most metals. The sequence of accelerations and collisions is continually repeated as the electrons move along the wire. The energy lost by the electrons in collisions increases the vibrations of the lattice, and the wire’s temperature rises. This temperature rise is just the familiar Joule heating of a wire by a current.

The argument just given does *not* apply to electrons in the full 1s band since there are no vacant states to their immediate left. These electrons are immobilized by the complete lack of empty states to transfer into. Thus the electrons in a filled band do not contribute to conduction.

This example already illustrates our two most important conclusions: Solids with a partially filled band are conductors; those in which all occupied bands are 100% full are insulators. Our next two examples show that these simple rules must be applied with care.

Beryllium: A Conductor

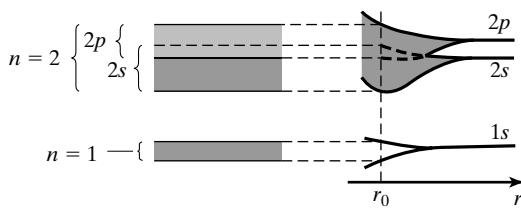
We turn now from lithium to the next element in the periodic table, beryllium. Its atomic configuration is $1s^22s^2$, and both of its occupied levels are filled. This might suggest that solid Be should have filled 1s and 2s bands and should therefore be an insulator, whereas Be is, in fact, a good conductor. The explanation of this apparent contradiction is not hard to find: As we assemble solid Be and its atomic levels fan out into bands, the 2s and 2p bands spread so far



424 Chapter 13 • Solids — Theory

FIGURE 13.20

The band structure of solid beryllium. As we assemble the solid from well-separated atoms, the $2s$ and $2p$ levels fan out so far that they overlap and form a single $n = 2$, valence band. Since this band could accommodate up to $8N$ electrons, it is less than half full, and the solid is a conductor.



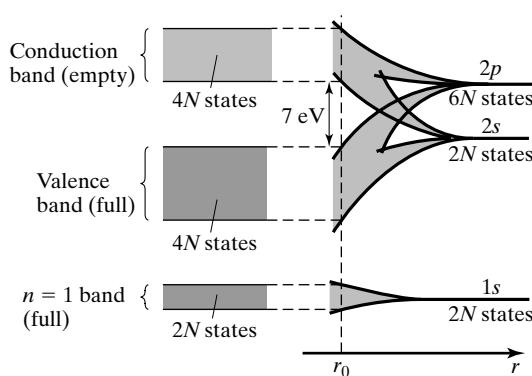
that they *overlap* one another, as shown in Figure 13.20. The $2s$ and $2p$ bands thus form a single valence band, which can accommodate altogether $8N$ electrons. (Remember an s band holds $2N$ electrons and a p band $6N$.) Thus, of the $4N$ electrons in the solid, $2N$ fill the $1s$ band, and the remaining $2N$ can only *partially* fill the valence band. Therefore, Be is a conductor.

Diamond: An Insulator

Atomic carbon has the configuration $1s^2 2s^2 2p^2$, with its $n = 2$, valence shell half-filled. This suggests that solid carbon should be a conductor, which, indeed it is in the form of graphite. However, solid carbon in the form of diamond is an outstanding *insulator*. To explain this puzzle, we must look carefully at the behavior of the energy levels as we assemble a diamond. In diamond the carbon atoms are covalently bonded. As we bring the atoms together and the covalent bonds form, the $2s$ and $2p$ levels fan out into bands, as we would expect. However, each level splits into *two separate bands*, as shown in Figure 13.21. (This is closely analogous to the development of the bonding and antibonding orbitals described in Section 12.4.) At the equilibrium separation r_0 , the lower two bands overlap each other, as do the upper two. Thus the band structure of diamond is as shown on the left of Figure 13.21. There is an $n = 1$ band that can hold $2N$ electrons and then two well-separated $n = 2$ bands, each of which can hold $4N$ electrons. Since carbon has $6N$ electrons in all, the lowest two bands are completely full, and the upper $n = 2$ band is completely empty. Because the lower $n = 2$ band holds all the valence electrons, it is called the valence band, and because electrons in the upper $n = 2$ band (if there were any) would contribute to conduction, this band is called the conduction band. With its valence band full and its conduction band empty, diamond is an insulator. It turns out that, as indicated in Figure 13.21, the gap between the valence and conduction bands of diamond is about 7 eV. Under normal conditions this means that no electrons can be excited into the empty conduction band. (At room temperature, for example, thermal motions could excite an electron, but

FIGURE 13.21

The band structure of diamond. Both the $2s$ and $2p$ levels split, each forming two separate bands. Each of the $2s$ subbands contains N states, and each of the $2p$ subbands contains $3N$. The lower $2s$ and the lower $2p$ bands then join to form a single band with $4N$ states in all, and likewise the upper $2s$ and upper $2p$.



the probability is so low that the expected number of excited electrons in a large diamond is far less than 1.) This makes diamond an extremely good insulator. This same large band gap also explains why diamond is transparent. Visible photons have energies of 2 to 3 eV and cannot excite any electrons across the 7 eV gap; therefore, visible light cannot be absorbed in diamond.

We mentioned in Section 13.2 that solid silicon (just below carbon in the periodic table) has the same crystal structure as diamond. The $3s$ and $3p$ levels in Si behave almost exactly like the $2s$ and $2p$ levels of diamond, and silicon, like diamond, is a nonconductor. There is, however, a subtle but crucial difference. The equilibrium separation r_0 in Si is larger than in diamond (0.24 nm in Si compared to 0.15 nm in diamond). It is easy to see in Figure 13.21 that a larger value of r_0 suggests a smaller band gap, and, indeed, the band gap in silicon is only about 1 eV (compared with 7 eV in diamond). Since visible photons can easily excite electrons across a 1 eV gap, light is quickly absorbed in silicon, which is therefore not transparent. The narrow band gap in silicon also has a dramatic effect on its conductivity. At low temperatures, all the valence electrons in Si are locked in the full valence band, and Si is an insulator. But as the temperature rises, a substantial number of electrons are excited across the narrow gap into the conduction band, and the conductivity increases rapidly. For this reason, silicon and other solids with a similar narrow gap between a full valence band and an empty conduction band are called **semiconductors**. Semiconductors include silicon and germanium, both from group IV of the periodic table, and several compounds of group III with group V, such as gallium arsenide. Semiconductors and their enormous technological importance are discussed in greater detail in Chapter 14.

13.6 The Drude Model of Conductivity

In this and the following two sections we look more closely at the conductivity of metals. We describe a simple but powerful model of electrical conduction, called the Drude* model. The Drude model is an example of a “semiclassical” model, meaning it is a hybrid of classical and quantum ideas. In the Drude model we think of conduction electrons as classical particles with a definite position and velocity, but to understand the interaction of these electrons with their surroundings, we use the results of quantum calculations. The classical view of electrons as point particles can be made roughly consistent with the quantum view of electrons as de-localized wave functions through the notion of wave packets, described in Section 6.7. A wave packet describing an electron in a metal can have both an average position and an average velocity, and so, within the constraints set by the Heisenberg uncertainty principle, we are allowed to view electrons in a metal as localized particles (see Problem 13.39).

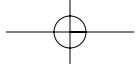
Before describing the Drude model, we begin with some definitions. The **electrical conductivity** σ of material is defined by the equation

$$j = \sigma E \quad (13.5)$$

Here E is the electric field[†] and j is the current density, which is defined as the current I per cross-sectional area A ; that is, $j = I/A$. Both the current density

*Pronounced “DREW-dah.”

[†] Here we use E for electric field, whereas, in earlier chapters, we used \mathcal{E} to avoid confusion with the symbol for energy.



426 Chapter 13 • Solids — Theory

and the electric field are vectors, but since we will be working in one dimension, we will drop the vector notation and use signs to represent direction, with (+) meaning rightward and (−) meaning leftward. In the SI system, conductivity has the units of $(\Omega \cdot m)^{-1}$, although one often sees the units of $(\Omega \cdot cm)^{-1}$ used. The **resistivity** ρ of a material is defined as the reciprocal of the conductivity, $\rho = 1/\sigma$, so resistivity has units of $\Omega \cdot m$ or $\Omega \cdot cm$. Equation (13.5) can then be rewritten as

$$E = j\rho \quad (13.6)$$

This equation is the microscopic version of Ohm's Law $V = IR$, which is discussed in Section 13.8. In Problem 13.40 you will show that (13.6) is indeed equivalent to Ohm's law.

A good electrical conductor, like copper, has a large σ and a small ρ , while a good electrical insulator, like glass, has an extremely small σ and a huge ρ . Table 13.2 and Figure 13.22 show the conductivities and resistivities of

TABLE 13.2

Electrical conductivities and resistivities of various materials.

Material	Conductivity ($\Omega \cdot m$) $^{-1}$	Resistivity ($\Omega \cdot m$)
<i>Metals (at 18°C)</i>		
Silver	6.21×10^7	1.61×10^{-8}
Copper	5.88×10^7	1.70×10^{-8}
Aluminum	3.65×10^7	2.74×10^{-8}
Lead	4.8×10^6	21.0×10^{-8}
Stainless steel	1.36×10^6	73×10^{-8}
<i>Ionic solutions (in water, 18°C)</i>		
1 molar* CuSO ₄	2560	3.9×10^{-4}
1 molar HCl	590	1.7×10^{-2}
Seawater (typical value)	4.5	0.22
Pure water	3.8×10^{-6}	2.6×10^5
<i>Semiconductors (pure, 20°C)</i>		
Germanium	2.2	0.45
Silicon	1.6×10^{-3}	640
<i>Insulators</i>		
Glass	$10^{-10} - 10^{-14}$	$10^{10} - 10^{14}$
Rubber	$10^{-13} - 10^{-16}$	$10^{13} - 10^{16}$

*1 molar means 1 mol of solute per liter of solution.

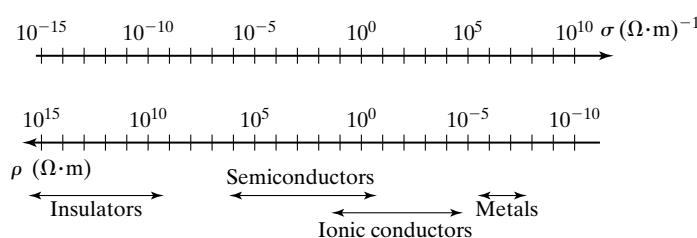


FIGURE 13.22

The range of conductivities and resistivities found in ordinary materials varies by many orders of magnitude.

Section 13.6 • The Drude Model of Conductivity **427**

various materials. Notice that both vary over some 24 orders of magnitude — an enormous range — from 10^7 or 10^8 $(\Omega \cdot m)^{-1}$ for good metals to 10^{-16} $(\Omega \cdot m)^{-1}$ for insulators.

In 1900, just 3 years after J. J. Thomson discovered the electron and well before the development of quantum mechanics, Paul Drude (German, 1863–1906) proposed a model of electrical conduction in metals, a model which is still used today because of its simplicity and usefulness. Drude made the bold assumption that the electrons in a metal behave as a gas of free particles. These free electrons, which we now understand are the conduction electrons, accelerate when an electric field is applied. If there were no frictional forces present, then the electrons would accelerate without limit, leading to an ever-increasing current. Since such runaway currents are never observed, there must be some mechanism at work that acts to slow the electrons. Drude thought that the electrons must collide with the ions, but this turns out to be incorrect. Because of a quantum-mechanical effect that we will describe later, the electrons do not collide with the ions. Instead they collide with both “impurity atoms” and “lattice vibrations.” These terms will be explained in detail in the next section.

Regardless of what the electrons collide with, we can define a collision time τ as an average time between collisions. Between collisions, an electron with charge $q = -e$ in an electric field E will experience a force $F = qE = -eE$ and hence an acceleration $a = F/m = -eE/m$, where m is the mass of the electron. If the electron starts from rest, then, in a time t , the electron will acquire a velocity

$$v = at = -\frac{eE}{m}t \quad (13.7)$$

Here, the minus sign indicates that for negatively charged electrons, the direction of the velocity is opposite the direction of the electric field. The effect of collisions, whatever their origin, is to momentarily interrupt this free acceleration and decrease, stop, or reverse the forward velocity. Our picture of the electron motion is then one of a very jerky start-stop kind of motion. In the competition between the acceleration due to the E field and the deceleration due to collisions, the electron acquires an average velocity called the **drift velocity**, v_d . With Eq. (13.7) in mind, we now *define* the collision time τ by the relation

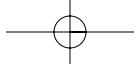
$$v_d = -\frac{eE}{m}\tau \quad (13.8)$$

The collision time is a representative average time between collisions. A given electron actually experiences a wide range of time intervals between successive collisions, and properly defining an “average” time interval has some surprising subtleties, which are explored in Problem 13.41.

Drude related this collision time τ to the electrical conductivity σ by the following argument. First, we show that the current density j can be written as

$$j = nqv_d \quad (13.9)$$

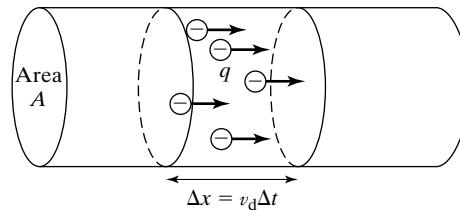
where n is the concentration (number per volume) of charge carriers, q is the charge per carrier, and v_d is the drift velocity. To understand this equation, consider a wire carrying a current I , as in Figure 13.23. In a time Δt , all the charges



428 Chapter 13 • Solids — Theory

FIGURE 13.23

A current-carrying wire contains n charges per volume, each with charge q and moving with average speed v_d .



will move a distance $\Delta x = v_d \Delta t$ as shown. A cylindrical portion of the wire of length Δx and cross-sectional area A contains a total charge

$$\begin{aligned}\Delta Q &= (\text{charge per carrier}) \times (\text{number of carriers}) \\ &= (\text{charge per carrier}) \times (\text{number per volume}) \times (\text{volume}) \\ &= qn(\Delta x A) = qn(v_d \Delta t A)\end{aligned}$$

All the charges making up this total charge ΔQ will pass through the right end of the cylinder during a time interval Δt , and the resulting current is $I = \Delta Q / \Delta t$. The current density is then

$$j = \frac{I}{A} = \frac{\Delta Q}{\Delta t A} = qn v_d$$

and we have proved (13.9).

Now inserting (13.8) into (13.9), and setting $q = -e$, we obtain

$$j = nq\left(\frac{-eE}{m}\tau\right) = \left(\frac{ne^2\tau}{m}\right)E \quad (13.10)$$

Comparing (13.10) with the equation $j = \sigma E$, we conclude that

$$\sigma = \frac{ne^2\tau}{m} \quad (13.11)$$

This equation is the central result of Drude's model of conductivity. It says that the conductivity is large when there are many charge carriers per volume (large n), few collisions (large τ), and light charge carriers (small m). The dependence on mass can be understood intuitively by noting that low-mass carriers accelerate more quickly in an applied field ($a = qE/m$), leading to a larger σ . Consequently, *ionic conductors* such as electrolytic solutions, in which massive ions carry the current, have a much smaller conductivity than normal metals, in which light electrons are the charge carriers, as seen in Table 13.2.

Example 13.3

A copper wire of diameter 1.0 mm (radius $r = 0.5$ mm) carries a current of $I = 10$ A, which is a fairly large current for such a wire. Estimate the magnitude of the drift velocity v_d .

According to (13.9), $j = nqv_d$, the drift velocity is easily computed if the current density j and the carrier concentration n are known. The current density is readily calculated as

$$j = \frac{I}{A} = \frac{I}{\pi r^2} = \frac{10 \text{ A}}{\pi(5 \times 10^{-4} \text{ m})^2} = 1.27 \times 10^7 \text{ A/m}^2 = 1270 \text{ A/cm}^2$$

(Here, we give the number in both SI units, as well in the more practical units of A/cm², often used by electrical engineers.) Computing n , the number density of conduction electrons in copper, is not quite so easy. One can compute it *roughly* by noting that for most solids, the nearest-neighbor atoms are about $r_0 = 0.3$ nm apart, and for copper, the valence or number of conduction electrons per atom is 1, as one might surmise by noting its electronic configuration — a single s electron outside a d shell. We conclude that the number density of charge carriers for copper is about

$$n \approx \frac{1}{r_0^3} = \frac{1}{(0.3 \times 10^{-9} \text{ m})^3} \approx 4 \times 10^{28} \text{ m}^{-3}$$

(The measured value is actually a factor of 2 higher than this.) Finally, we can compute the drift velocity as

$$v_d = \frac{j}{nq} = \frac{1.27 \times 10^7 \text{ A/m}^2}{(4 \times 10^{28} \text{ m}^{-3})(1.6 \times 10^{-19} \text{ C})} \approx 2 \times 10^{-3} \text{ m/s} \approx 2 \text{ mm/s}$$

This is a small speed; a given electron in this wire takes more than 8 minutes to travel a meter. When you flip a light switch, the electromagnetic signal travels along the wire at nearly the speed of light, but the electrons themselves move quite slowly. A long garden hose filled with water behaves similarly. When you open the faucet, water comes out the far end of the hose immediately, due to the incompressibility of water, while a given water molecule takes several seconds to travel the length of the hose.

13.7 Electron Collisions in Metals

We now return to the nature of the collisions experienced by conduction electrons in a metal. Understanding the origin of these collisions requires a knowledge of quantum mechanics, which did not exist in Drude's time.

In a metal, the atomic nuclei surrounded by their inner core electrons form an array of positively charged ion cores. A conduction electron sees these ion cores as an array of attractive potential wells separated by barriers, as shown in Figure 13.24. To understand how an electron interacts with these barriers, we must recall some qualitative features of quantum-mechanical tunneling, which was discussed in Chapter 7.

When a free electron wave packet encounters a single potential barrier, as in Figure 13.25(a), there is, in general, a reflected wave and a transmitted wave. If the barrier height U_0 is high enough or the barrier width L is sufficiently large, the electron will be almost entirely reflected. If there are two barriers, as in Figure 13.25(b), a remarkable phenomenon called *resonant tunneling* can occur. If the energy of the incident electron wave has certain

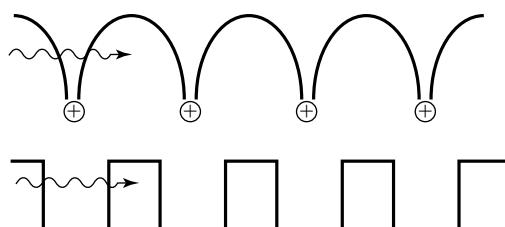
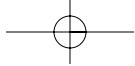


FIGURE 13.24

The potential energy $U(x)$ due to the ion cores in a metal is qualitatively similar to a periodic array of square barriers.



430 Chapter 13 • Solids — Theory

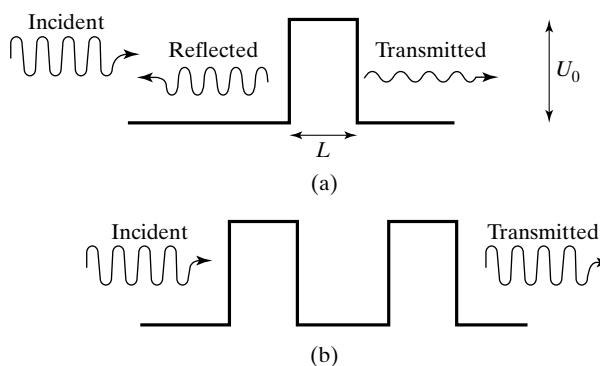


FIGURE 13.25

(a) An electron wave packet incident on a single barrier is partially transmitted, partially reflected. (b) If an electron incident on a double barrier has the correct energy for resonant transmission, it will be fully transmitted.

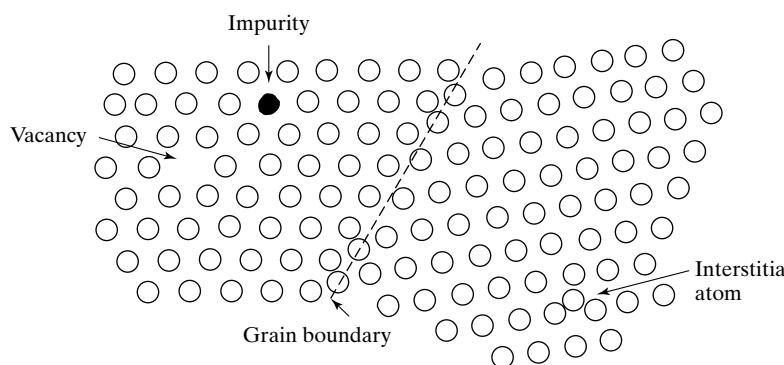


FIGURE 13.26

Various types of defects appearing in a crystalline solid.

special values, then the waves reflected from the two barriers will interfere destructively, resulting in no net reflected wave and 100% transmission of the incident wave. There is an analogous process in the field of optics: an antireflection coating on a glass surface. If a wave of the correct frequency is incident on a coating of given thickness and refractive index, it experiences 100% transmission.

If instead of two barriers, there is an infinite array of uniformly spaced identical barriers, as in Figure 13.24, it turns out that resonant transmission occurs over a range of electron energies. In this band of energies, the electron behaves essentially as a free electron. It can travel forever, without scattering. This range of energies in which free propagation occurs is precisely what we called the conduction band in Section 13.5.

If the nuclei in a metal formed a *perfectly* regular array, then, because of resonant tunneling, the conduction electrons could zip right through the metal unimpeded, and the metal would have infinite conductivity. However, any disruption in the periodicity of the lattice causes electron scattering, and there are two causes of nonperiodicity that are always present in a metal. First, any metal sample always contains some impurities, no matter how carefully it has been purified. Impurities may be foreign atoms, or they may be *defects*, such as vacancies (missing atoms), interstitial atoms, or grain boundaries, all of which are illustrated in Figure 13.26.

Besides impurities and other defects, *lattice vibrations* also disrupt the periodicity of the lattice and cause electron scattering. At any nonzero temperature, the atoms of a solid vibrate about their equilibrium positions, as in the ball-and-spring model of a solid shown in Figure 13.27. The higher the temperature, the more vigorous the vibrations. At a high enough temperature, this

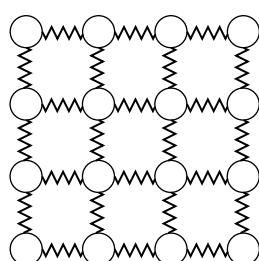
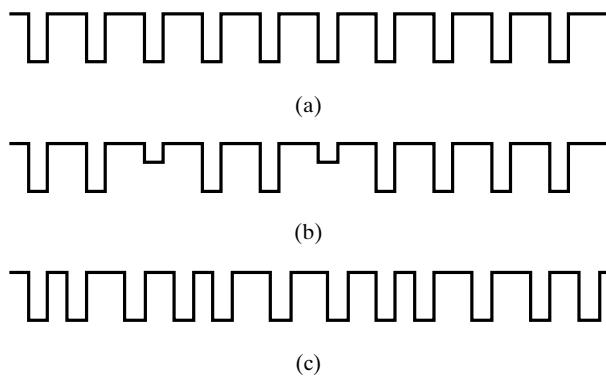


FIGURE 13.27

Ball-and-spring model of an atomic lattice.

**FIGURE 13.28**

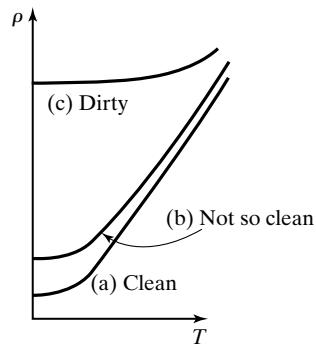
(a) A perfectly periodic potential such as found in a defect-free metal at low temperatures. (b) A potential with some impurity sites. (c) A nonperiodic potential due to lattice vibrations as found in a metal at high temperature.

“jiggling” of the atoms becomes so great that the solid melts. In the correct quantum-mechanical description of this ball-and-spring model of a solid, this vibrational energy is quantized and the quanta of energy are called “phonons.” In classical language, higher temperature means larger amplitude vibrations, while in the language of quantum mechanics, higher temperature means more phonons and, consequently, more **electron-phonon scattering**. It turns out that at room temperature and above ($T \gtrsim 300$ K), the rate of electron-phonon scattering is proportional to temperature, leading to a resistivity proportional to temperature.

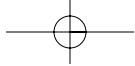
In Figure 13.28, we indicate schematically the potential that a conduction electron encounters in three environments: (a) a perfectly periodic potential such as is almost achieved in very pure metals at very low temperatures; (b) the nonperiodic potential seen in a metal with some impurities, in which case different types of atoms produce different well depths; and (c) the nonperiodic potential found in a pure sample at high temperatures, in which case lattice vibrations produce a nonuniform spacing of the atoms. In cases (b) and (c), compared with case (a), there is much more scattering of electrons, a shorter time τ between collisions, and a smaller conductivity $\sigma = ne^2\tau/m$.

The general shape of the curves of resistivity versus temperature found in metal samples are displayed Figure 13.29. Three curves are shown corresponding to three different samples:

- (a) **A “clean” sample, such as nearly pure copper or aluminum.** In this case, at higher temperatures (room temperature and above), lattice vibrations are the dominant source of electron scattering leading to a linear dependence of ρ on T . At very low temperatures (a few tens of Kelvin or lower) lattice vibrations are so weak that the impurities (although few in number) dominate electron scattering. This leads to a temperature-independent resistivity, as indicated by the short horizontal section of the curve near $T = 0$.
- (b) **An “ordinary” sample, very like (a) but not so pure.** Here, the impurity scattering is larger than in case (a), leading to a higher resistivity at low T . The dependence of the low-temperature resistivity on purity leads to a simple way to quantify the purity of samples: The residual resistivity ratio (RRR) or “triple-R” is defined as the ratio of the resistivity at room temperature to that at helium temperatures (a few kelvins). The higher the triple-R, the more pure the sample. Carefully purified metal samples have a triple-R of a few hundred; very dirty samples have a triple-R of less than 2.
- (c) **A “dirty” (very impure) metal or an alloy.** An alloy is a mixture of metals; an example is brass, which is a mixture of copper and tin. In an alloy,

**FIGURE 13.29**

Resistivity ρ versus temperature T for three metal samples: (a) a nearly pure metal sample, (b) a less pure sample, and (c) a very dirty metal sample or an alloy.



432 Chapter 13 • Solids — Theory

almost every atom can be regarded as an “impurity” and, consequently, the impurity scattering is so large that it dominates the resistivity at all but the highest temperatures. Another example in this category is stainless steel, which is iron with a very high concentration of other elements including chromium and nickel. Stainless steel can be regarded as an alloy or as very “dirty” iron.

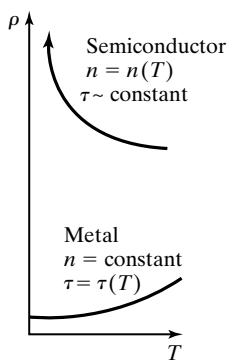


FIGURE 13.30

Typical temperature dependence of the resistivity of a semiconductor and a metal.

In Figure 13.30 we plot the resistivity ρ against temperature T for a typical metal, such as copper, and for a pure semiconductor, such as silicon. Drude’s equation $\sigma = ne^2\tau/m$ gives us insight into the very different behaviors of metals and semiconductors seen in this graph. In metals the temperature dependence of the conductivity is due entirely to the temperature dependence of the collision time τ . The carrier concentration n is temperature-independent, while the collision time τ decreases with increasing temperature. As the temperature increases, τ decreases because of increased scattering from lattice vibrations; consequently, the conductivity decreases and the resistivity rises. In semiconductors, however, the temperature dependence of the resistivity is due mostly to the temperature dependence of the carrier concentration. In semiconductors, n is small compared with metals and very strongly temperature-dependent. The smallness of n results in a large resistivity compared with metals. In pure (undoped) semiconductors, the carrier concentration n is thermally activated: As the temperature is increased, more and more electrons are thermally excited from the valence band into the conduction band, across the band gap; n increases rapidly with increasing temperature, leading to a decrease in resistivity.

13.8 The Fermi Speed

A very important feature of the conductivity of metals, a feature we have not yet discussed, is that the conductivity σ in the equation $j = \sigma E$ is *independent* of the applied electric field E or the current density j . That is to say, although j and E may vary considerably, the *ratio* j/E remains constant. Conductors that exhibit a field-independent conductivity are called **ohmic** conductors. Resistors made from ohmic conductors obey **Ohm’s law**, which is the statement that $V/I = R = \text{constant}$, where I is the current through the sample, V is voltage across it, and R is its resistance. It is important to note that the equation $R = V/I$, by itself, is not Ohm’s law; it is merely the definition of the resistance R of a sample. Ohm’s law is the statement that the resistance (defined as V/I) is a *constant*, independent of V or I . Most conductors are ohmic, so long as the current density j is low enough that Joule heating does not raise the temperature of the sample. If significant Joule heating does occur (as in most incandescent lightbulbs), the increased temperature causes an increase in ρ because of increased electron-phonon scattering, and the sample under study exhibits non-ohmic behavior.

The fact that the conductivity of a metal is independent of the applied electric field E is surprising for the following reason: The conductivity depends on the collision time τ according to the Drude formula (13.11), $\sigma = ne^2\tau/m$. At first glance, it might seem that the collision time should depend on the field E . After all, if the field E is increased, the acceleration a of a conduction electron will increase (according to $a = qE/m$); and if the electron accelerates more quickly, it will surely encounter an impurity or phonon more quickly, resulting in a shorter collision time τ and a smaller conductivity σ . This argument,

Section 13.8 • The Fermi Speed **433**

which implies that σ depends on E , would be correct if conduction electrons had zero initial velocity after a collision. However, because of the Pauli principle, as we will explain shortly, the electrons in a metal always have a very large intrinsic speed, called the **Fermi speed** v_F . The Fermi speed is about 0.3% of the speed of light ($v_F \approx 0.003c \approx 10^6$ m/s). In contrast, the magnitude of the drift velocity v_d (the mean speed due to the applied field E), is minuscule ($v_d \approx 2$ mm/s, as in Example 13.3). Consequently, the increase in speed due to the E field (and hence the decrease in time until the next collision) is completely negligible compared to the initial huge Fermi speed of the conduction electron, and the collision time is almost completely independent of the field E . The fact that τ is independent of E is the origin of Ohm's law.

The motion of conduction electrons is thus one of extremely rapid, random motion due to the large Fermi speed and, superimposed on this, a very small bias toward one direction due to the tiny drift velocity. As shown in Figure 13.31, a conduction electron in a metal follows a random zigzag path even with zero applied field ($E = 0$), regardless of the temperature. This is an example of **random-walk diffusive motion**, which was described in Section 3.8. The **mean free path*** l is defined as the mean distance an electron travels between collisions, that is, the step length in the random walk. The mean free path is related to the Fermi speed and the collision time τ by

$$v_F = \frac{l}{\tau} \quad (13.12)$$

The mean free path (mfp) depends on the purity and temperature of the sample. Very dirty samples, such as alloys, or metals at very high temperatures, may have an mfp of nanometers or less. Metals of ordinary purity at room temperature have an mfp of roughly 10–100 nm. While high-purity (clean) samples at low temperatures may have an mfp of many micrometers or even millimeters.

Example 13.4

Using the Drude model and the known Fermi speed, estimate the collision time and the mean free path for conduction electrons in copper at room temperature.

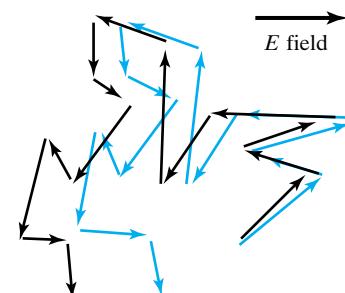
Drude's equation, $\sigma = ne^2\tau/m$, allows us to compute the collision time, τ , from the measured conductivity, σ , given in Table 13.2. From Example 13.3, the conduction electron density is about $n = 8 \times 10^{28}$ m⁻³. Thus,

$$\tau = \frac{m\sigma}{ne^2} = \frac{(9.11 \times 10^{-31} \text{ kg}) \times [5.9 \times 10^7 (\Omega \cdot \text{m})^{-1}]}{(8 \times 10^{28} \text{ m}^{-3}) \times (1.6 \times 10^{-19} \text{ C})^2} \approx 3 \times 10^{-14} \text{ s}$$

Assuming that the Fermi speed is approximately 10⁶ m/s, we have for the mfp

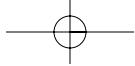
$$l = v_F\tau \approx (10^6 \text{ m/s}) \times (3 \times 10^{-14} \text{ s}) = 3 \times 10^{-8} \text{ m} = 30 \text{ nm}$$

This is approximately 100 *lattice constants*. (A lattice constant is defined as the linear size of a “unit cell” in a crystal; it is roughly equal to the distance between nearest-neighbor atoms.)

**FIGURE 13.31**

The motion of a conduction electron in a metal is a random walk. In the presence of an applied electric field, there is a very slight drift antiparallel to the field. The black path is the trajectory of an electron in the presence of an electric field; the light gray path is the trajectory that the electron would have followed if the E field was zero. The drift is greatly exaggerated here. For typical current densities of 1 A/cm², the actual bias is about 1 part in 10¹⁰.

*In Section 3.8 we used the symbol λ for the mean free path. Here, we use the symbol l , to avoid confusion with the de Broglie wavelength.



434 Chapter 13 • Solids — Theory

Lev Landau
(1908–1968, Russian)



One of the most influential theorists and teachers of the twentieth century, Landau worked in many areas of physics. His contributions to solid-state physics included theories of magnetism, superconductivity, and phase transitions. He won the Nobel Prize for physics in 1962, but was unable to attend the ceremony because he was seriously injured in a car accident. He suffered brain damage and never fully recovered.

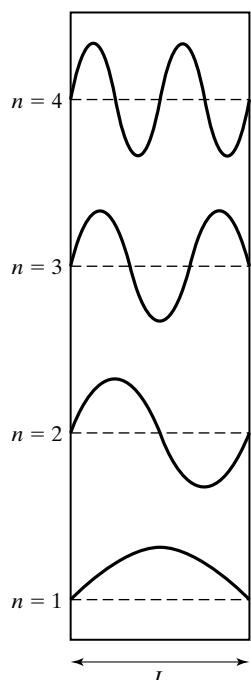


FIGURE 13.32

Wave functions of a particle in a 1-D box of length L . (The vertical axis representing energies is not to scale.)

We now consider why conduction electrons in metals have such a large intrinsic speed, the Fermi speed. We first note that the Fermi speed is not simply the *thermal* speed of the electrons. As described in Section 3.7, in a *classical* gas of particles of mass m at temperature T , the average **thermal speed** v of the particles is given by the expression $\frac{1}{2}mv^2 = \frac{3}{2}kT$. For electrons at room temperature, this thermal speed is about 10^5 m/s, a factor of 10 smaller than the Fermi speed. Another difference between the thermal and Fermi speeds is that the Fermi speed is independent of temperature, while the thermal speed rises with increasing temperature.

The large Fermi speed of conduction electrons is a consequence of the Pauli exclusion principle. To see why, we begin by recalling that conduction electrons in a metal can be regarded as free electrons in a box, a problem that was considered in Chapters 7 and 8. Of course, the electrons are not truly free; that is, they are not in a region where the potential energy function $U(x)$ is constant. An individual electron experiences the attraction of the ion cores of the atoms in the sample, as well as a repulsive potential energy due to the other conduction electrons. However, the Coulomb repulsion from the other electrons nearly cancels the Coulomb attraction to the ion cores, and the electrons behave as “nearly free” particles. The validity of this nearly-free-electron approximation was rigorously demonstrated by the Russian theoretical physicist Lev Landau.

Let us imagine placing electrons in an empty box, one at a time. Fig. 13.32 shows the wave functions of the various electron states for the case of a one-dimensional box of length L . The first two electrons will occupy the ground state ($n = 1$) with one electron spin up and the other spin down. Because of the Pauli exclusion principle, the next two electrons will be forced to occupy the first excited state ($n = 2$); the next two will occupy the second excited state ($n = 3$), etc. As more electrons are poured into the box, they are forced to occupy higher and higher energy states. This filling of states is the partial filling of conduction band states described in Section 13.5. If there are very many electrons in the box (many electrons in the conduction band), the average energy of an electron will be quite large compared to thermal energies. This is the reason that conduction electrons in a metal are so energetic. In a metal the density of conduction electrons is such that all of the quantum states up to a high energy level are occupied. As we said in Section 13.5, this highest occupied level is the Fermi level, and its energy is called the Fermi energy $E_F = \frac{1}{2}mv_F^2$. (As you might guess, all this was first explained by the Italian-American physicist Enrico Fermi.) In a metal the Fermi energy has a value of a few electron volts, and this corresponds to a speed, the Fermi speed, of about 10^6 m/s.

We can estimate the value of the Fermi energy by the following argument: Notice in Figure 13.32, that the n^{th} state of the system has a wavelength λ , which satisfies

$$n \frac{\lambda}{2} = L \quad (13.13)$$

When there are N electrons in the system, the highest occupied state is the state with $n = N/2$, so we can write

$$N \frac{\lambda}{4} = L \quad (13.14)$$

If we temporarily regard electrons as classical (rather than quantum) objects, then the average distance between nearest-neighbor electrons is $r = L/N$,

Section 13.8 • The Fermi Speed **435**

which by Eq. (13.14) becomes $r = \lambda/4$. Although we have made this argument in one dimension (1-D), a similar result is true in 3-D; that is, in either 1-D or 3-D the average distance between electrons is very roughly equal to the wavelength of the highest occupied state. In 3-D a careful treatment gives $r \approx \lambda/2$, rather than the 1-D result, $r = \lambda/4$. We can now compute the kinetic energy of electrons at the Fermi energy. We write

$$E = \frac{p^2}{2m} = \frac{h^2}{2m\lambda^2} \approx \frac{h^2}{2m(2r)^2} = \frac{h^2}{8mr^2} \quad (13.15)$$

where we have taken $r \approx \lambda/2$ as the nearest-neighbor distance. If we consider a monovalent metal, that is, a metal with one conduction electron per atom, then the average distance r between conduction electrons is the same as the average distance between atoms, which is about 0.3 nm for most solids. We then have for the Fermi energy,

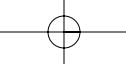
$$\begin{aligned} E_F &\approx \frac{h^2}{8mr^2} = \frac{(6.6 \times 10^{-34} \text{ J} \cdot \text{s})^2}{8 \times (9.1 \times 10^{-31} \text{ kg}) \times (0.3 \times 10^{-9} \text{ m})^2} \\ &\approx 7 \times 10^{-19} \text{ J} \approx 4 \text{ eV} \end{aligned} \quad (13.16)$$

The Fermi energy varies from metal to metal because of variations in the valence of the metal and the density of the atoms, but it always falls in the range from 2 to 12 eV. This is the same order of magnitude as the energy of chemical bonds, so the result is not too surprising. We can now, finally, compute the Fermi speed. Since $E_F = \frac{1}{2}mv_F^2$,

$$v_F = \sqrt{2E_F/m} \approx \sqrt{2 \times (7 \times 10^{-19} \text{ J})/(9.1 \times 10^{-31} \text{ kg})} = 1.2 \times 10^6 \text{ m/s} \quad (13.17)$$

Note, once again, how enormous this speed is, compared with the drift speed.

We conclude our discussion of the Drude model with the story of a curious historical accident. As we have mentioned before, Drude's work was done well before the advent of quantum mechanics, so there was no way for Drude to compute the Fermi speed or the mean free path. Why then was Drude's model immediately successful? It was because Drude's calculations contained two large mistakes that canceled! From the known values of the conductivity of metals, Drude used his formula (13.11) to correctly compute the collision time τ . As a check on the theory, he also computed the collision time using $\tau = l/v$, where l is the mean free path and v is the mean electron speed. But his values for l and v were *both* too small by an order of magnitude. He assumed, incorrectly, that the electrons scattered from individual ions, so his estimate of the mean free path l was a few nanometers — smaller by a factor of at least 10 than the correct value for metals at room temperature. He also incorrectly assumed that the speed of the electrons is given by the *thermal* speed of about 10^5 m/s, which is about 10 times smaller than the correct Fermi speed. Because both numbers were wrong by about the same factor, the ratio, τ , was correct. Consequently, the model appeared to be nicely self-consistent and it was immediately accepted by physicists. More than 30 years passed before enough quantum mechanics was known for the correct picture to emerge. Today we remember Drude because his simple but wrong model leads to the correct formula, $\sigma = ne^2\tau/m$. In science, unlike politics, the good that physicists do lives after them, while the bad is oft interred with their bones.



436 Chapter 13 • Solids — Theory

13.9 Degeneracy Pressure*

*This material will be used in Section 13.10 but will not be needed after that.

In the previous sections we have described the properties of conduction electrons in metals. Because these electrons behave as if they were nearly free, they are often referred to as a “gas” of electrons. However, the behavior of this gas is quite unlike that of an ordinary gas of molecules, such as air. The electron gas is an example of a **degenerate Fermi gas** — called *degenerate* because the system is near its ground state,* and *Fermi* because it consists of *fermions*, that is, particles that obey the Pauli exclusion principle. One of the properties of a degenerate Fermi gas is that it is extremely hard to compress. If you squeeze on a block of metal, like copper, you find that to compress it significantly requires pressures of order a million atmospheres. This incompressibility of metals is largely explained by a quantum-mechanical effect called **degeneracy pressure**, which occurs in degenerate Fermi gases.

The basic physics of degeneracy pressure can be understood by considering the simplest possible Fermi gas, namely a single electron, confined to a rigid 1-D box and in its ground state. The ground-state wave function is shown in Figure 13.33. If the box has length L , the wavelength of the ground state is $\lambda = 2L$ and the ground-state energy is, from (13.15), $E = h^2/(8mL^2)$. If the box is compressed, decreasing L , then the kinetic energy of the ground state must increase since E is proportional to $1/L^2$. (Recall that in quantum mechanics, shorter wavelength corresponds to higher kinetic energy.) So compressing the container and decreasing the wavelength necessarily increases the energy. Since the energy is increased, work must be done to perform the compression; the gas is hard to compress. The electron wants to lower its energy by spreading out over a larger volume; that is, it wants to push the walls of the enclosure outward. This pressure that the confined electron exerts on the walls of its container is called the degeneracy pressure.

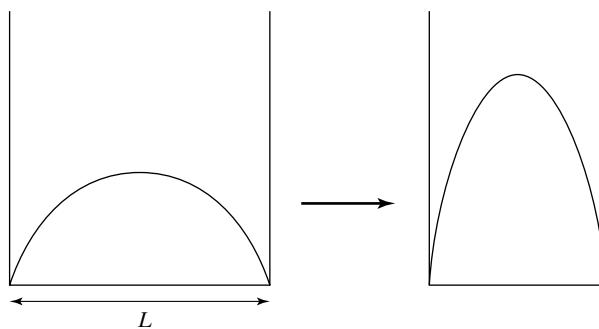


FIGURE 13.33

The ground-state energy E of a quantum particle in a rigid box depends on the size L of the box as $E \propto L^{-2}$. If the box is compressed, the energy of the ground state increases. This is the origin of degeneracy pressure.

*There are two distinct uses of the term “degenerate” in quantum mechanics. The first use was described in Section 8.3: Two wave functions are degenerate if they have the same energy. Now we encounter a different use: A many-particle system is degenerate if the system is near its ground state so that its particles are “crowded” into a limited number of single-particle states; the system is nondegenerate if the temperature is high enough that the system is far from its ground state and its particles have many empty single-particle states to choose from.

Section 13.9 • Degeneracy Pressure 437

This argument is readily extended to many electrons (or fermions) in a 3-D box. The n^{th} quantum state of an electron in a 1-D rigid box has wavelength λ such that $n\lambda/2 = L$, and its energy is

$$E_n = \frac{p^2}{2m} = \frac{h^2}{2m\lambda^2} = \frac{h^2}{8mL^2} n^2 \quad (13.18)$$

In a 3-D cubical box of volume $V = L^3$, the energy becomes

$$E = \frac{h^2}{8mL^2} (n_x^2 + n_y^2 + n_z^2) \quad (13.19)$$

where n_x , n_y , and n_z are the quantum numbers labeling the state. For many noninteracting electrons in a 3-D box, the total energy is the sum of the one-electron energies

$$E_{\text{tot}} = \frac{h^2}{8mL^2} \sum_i (n_{xi}^2 + n_{yi}^2 + n_{zi}^2) \quad (13.20)$$

where the sum is over all occupied states i . At low temperatures, each level up to the Fermi level will be occupied by 2 electrons. Now a marvelous simplification occurs. In our present analysis, only the dependence of the energy on the dimensions of the container matters. So we rewrite the total energy (13.20) as

$$E_{\text{tot}} = \frac{A}{L^2} = \frac{A}{V^{2/3}} \quad (13.21)$$

where A is a constant and $V = L^3$ is the volume of the cubical box.

We can use this simple expression for the total energy to compute the pressure of the Fermi gas. Recall from thermodynamics that for a gas of volume V and pressure p , if the volume of the gas is changed by an amount dV , and no heat is added or removed (a so-called *adiabatic* change), then the work done on the gas is equal to the change in energy

$$dW = dE_{\text{tot}} = -p dV \quad (13.22)$$

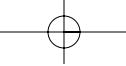
The minus sign here indicates that the energy increases when the volume decreases. The pressure exerted by the gas on the walls of its container is thus

$$p = -\frac{dE_{\text{tot}}}{dV} \quad (\text{adiabatic conditions}) \quad (13.23)$$

Combining (13.21) and (13.23), we find that the degeneracy pressure is

$$p = -\frac{d}{dV} \left(\frac{A}{V^{2/3}} \right) = \frac{2}{3} AV^{-5/3} = \frac{2}{3} AV^{-2/3} V^{-1} = \frac{2}{3} \frac{E_{\text{tot}}}{V} \quad (13.24)$$

where in the last step we have used (13.21) again.



438 Chapter 13 • Solids — Theory

Example 13.5

Estimate the degeneracy pressure of an electron gas with a concentration equal to that in an ordinary metal.

We use Eq. (13.24) and our earlier estimate (13.16) of the Fermi energy, $E_F \approx 4$ eV. For a degenerate Fermi gas, the particles occupy single-particle states with energies that range from zero to E_F . We therefore expect that the average energy per particle is roughly $E_F/2$ [the exact answer turns out to be $(3/5)E_F$]. The total energy of a gas is the sum of the energies of all the particles in the gas. So, if there are N particles in the gas, the total energy is $E_{\text{tot}} \approx NE_F/2$, and Eq. (13.24) becomes

$$p = \frac{2}{3} \frac{E_{\text{tot}}}{V} \approx \frac{1}{3} \frac{NE_F}{V} = \frac{1}{3} nE_F \quad (13.25)$$

where $n = N/V$ is the concentration of electrons (number/volume). For a metal, this is roughly $n = 1/a^3$, where $a \approx 0.3$ nm is a typical lattice constant, and so the pressure is

$$p \approx \frac{E_F}{3a^3} \approx \frac{(4 \text{ eV}) \times (1.6 \times 10^{-19} \text{ J/eV})}{(3) \times (0.3 \times 10^{-9} \text{ m})^3} \approx 10^{11} \text{ N/m}^2 \approx 10^6 \text{ atm} \quad (13.26)$$

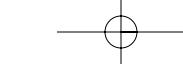
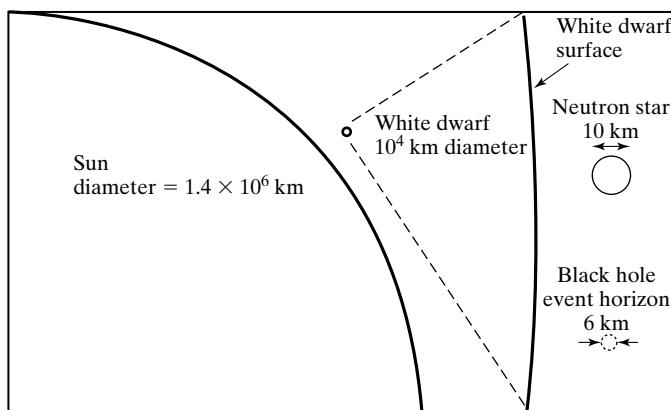
(In the last step we used 1 atmosphere = 1 atm $\approx 10^5$ N/m².) This is an enormous pressure, comparable to the pressure at the center of the earth. The reason metals do not blow apart is because this pressure is balanced by the very strong Coulomb attraction between the negative electrons and the positive ion cores.

Degeneracy pressure makes degenerate Fermi gases much harder to compress than ordinary molecular gases. If you squeeze on a molecular gas, such as air, at STP (standard temperature and pressure: $T = 273$ K, $p = 1$ atm) with a pressure of two atmospheres, its volume will decrease by about a factor of two. This is easy to deduce from the ideal gas law. In contrast, to substantially compress a solid such as a block of copper requires a pressure of the same order of magnitude as the degeneracy pressure — about 10^6 atm. (See Problem 13.51.). Degeneracy pressure has some remarkable consequences for the evolution of stars, as we will see in the next section.

13.10 White Dwarfs, Neutron Stars, and Black Holes*

*This section, though it describes some of the most remarkable discoveries of twentieth-century astronomy, contains material that will not be used again and can be omitted without loss of continuity.

The most spectacular examples of degenerate Fermi gases occur in the realm of astrophysics. Degeneracy pressure profoundly influences the evolution of stars and leads to fantastically dense stellar remnants called white dwarfs and neutron stars. An ordinary star, like the sun, has a long, stable “middle age” during which the star shines because the temperature and pressure near its core are high enough for hydrogen nuclei (protons) to fuse to form helium. This fusion reaction produces the energy that makes the star shine, and it also produces an outward pressure that counteracts the inward pull of gravity, so

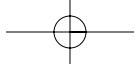
Section 13.10 • White Dwarfs, Neutron Stars, and Black Holes **439****FIGURE 13.34**

Relative sizes of the sun, a white dwarf star, a neutron star, and a black hole, each with one solar mass. Black holes resulting from stellar collapse are believed to have masses greater than 5 solar masses, and so are larger than the 1-solar-mass black hole shown here.

the star has a stable size. Eventually, however, the supply of hydrogen near the star's core is depleted to the extent that the fusion reaction slows and the core begins to collapse because the inward pull of gravity is unopposed. Details of the subsequent evolution depend on the mass of the star, but the story is usually violent. As the core collapses, it becomes hot enough that helium begins to fuse to form even heavier elements, and the outer layers of the star are blown off by the violent reactions in the core. Eventually, the nuclear fuel is exhausted and the star further collapses under its own gravity. If the star is not too much more massive than the sun, it will collapse until the electron degeneracy pressure (13.24) grows large enough to counteract gravity. Such a collapsed star, called a **white dwarf**, is extremely dense, with a mass comparable to the sun's mass packed into a sphere the size of the earth, as in Figure 13.34. Over time, the star cools by radiation and becomes a dark, cold, cinder. Countless white dwarf stars have been identified by astronomers; they are common objects in the sky.

In 1937 a young theoretical physicist, S. Chandrasekhar, discovered that if a star's mass is greater than 1.4 solar masses — the so-called **Chandrasekhar limit** — even electron degeneracy pressure is insufficient to stop the gravitational collapse. As the star collapses into an object smaller than a white dwarf, a process called **inverse beta decay** occurs; at sufficiently high pressures, an electron will burrow into a proton, forming a neutron and a neutrino: $p + e^- \rightarrow n + \nu$. (See Section 17.4, for discussion of closely related reactions.) The star is then composed primarily of neutrons. Neutrons, like electrons and protons, are fermions; they obey the Pauli exclusion principle and can exert degeneracy pressure just as electrons do. If the star's mass is less than about 5 solar masses, neutron degeneracy pressure will balance gravity when the diameter of the star is about 10 km. This fantastically dense object is called a **neutron star**; the density at its center is about 10^{18} kg/m³, which is the density of “nuclear matter” in the interior of atomic nuclei. The escape speed at the surface of a neutron star is about $0.8c$ (where c is the speed of light). Problem 13.56 explores the calculation of the diameter of a neutron star.

Another remarkable feature of neutron stars is their rapid rate of rotation. As a star collapses, it must spin faster, to conserve angular momentum — just as skaters spin faster as they pull in their arms. Sun-sized stars typically have rotational periods of days to months; after collapse into an object of 10-km diameter, a neutron star spins at rates of 10 to 1000 times *per second*. The rapid rotation of the neutron star and its magnetic field interacts with nearby interstellar dust and gas to produce a bright light source that rotates with the



440 Chapter 13 • Solids — Theory

**Subrahmanyan
Chandrasekhar
(pronounced
Subra-MON-yon
Chandra-SAY-kar)**
(1910–1995, Indian-American)



Known simply as “Chandra” by friends and colleagues, Chandrasekhar was the foremost theoretical astrophysicist of the twentieth century. As a 20-year-old physics student on the voyage from his native India to graduate school in Cambridge, England, Chandra combined relativity with the new theory of quantum mechanics to predict the properties of white dwarf stars. The bizarre predictions of his theory originally met with skepticism, even ridicule, from astronomers. Among the critics was Chandra’s own thesis advisor, Sir Arthur Eddington. Undeterred, Chandra went on to enjoy a stellar career at the University of Chicago. In 1985 he received the Nobel Prize in physics for his work on stellar structures, begun 55 years earlier. NASA’s Chandra X-ray Observatory, an orbiting satellite, was named in his honor.



FIGURE 13.35

A probable black hole. The image on the right, taken by the Hubble space telescope, shows a torus-shaped cloud of glowing gas in rapid rotation at the center of a distant galaxy — the galaxy is seen on the left in an image taken from a ground-based telescope. The cloud is 3700 light-years across. The gravitating object at the center of the cloud has a mass of 300 million suns, as deduced from the size and rotation rate of the cloud.

star, causing the star to appear to flash on and off as seen from the earth. This rapidly flashing neutron star is called a **pulsar**, and since the 1960s several dozen pulsars have been positively identified.

If the mass of the collapsing star is greater than roughly 5 solar masses, even neutron degeneracy pressure is insufficient to counteract the inward pull of gravity and the star collapses into an *infinitely* dense, *infinitesimally* small object called a **singularity**. At present we have no understanding of physics in the vicinity of a singularity; all our theories (relativistic quantum mechanics, general relativity) break down under such conditions. Surrounding the singularity is an imaginary sphere called the **event horizon** that marks the boundary at which the escape speed equals the speed of light. The event horizon around a 10-solar-mass singularity is 60 km in diameter. No light and no signal of any kind can escape from inside the event horizon; things can fall in, but nothing can come out. The singularity and its surrounding event horizon are therefore called a **black hole**. The evidence for the existence of black holes is indirect but very strong. We can “see” black holes by the effect of their gravity on their surroundings. It is believed that supermassive black holes, of some 10^6 solar masses, reside at the centers of most galaxies, including our own Milky Way. See Figure 13.35.

13.11 Classical and Quantum Gases*

*This material will be needed in the following section and in Chapter 15, Statistical Mechanics, but not elsewhere in this text.

The most general definition of a “gas of particles” that one can give — a definition sufficiently broad that it is valid in the framework of either classical or quantum mechanics — is this: A gas is a collection of noninteracting or weakly interacting particles. A **classical gas** is one that is well described with classical

Section 13.11 • Classical and Quantum Gases **441**

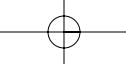
mechanics, while a **quantum gas** is one in which quantum effects are important. We have already seen in Section 13.9 one example of a quantum gas, namely, the degenerate Fermi gas formed by conduction electrons in a metal. We will see that a gas behaves classically when conditions are such that the particles of the gas are far apart, compared to their wave-packet diameters. However, for a gas of identical particles, if the particles overlap, quantum effects cannot be ignored. The properties of a quantum gas depend crucially on whether the particles of the gas are fermions or bosons. Before describing quantum gases, let us review the properties of fermions and bosons, which were briefly discussed in Section 10.5.

All particles can be classified as either fermions or bosons. A fermion is defined as a particle that obeys the Pauli exclusion principle. Examples of fermions are the electron, the proton, and the neutron. In Section 10.4, the Pauli exclusion principle was stated as “No two electrons can occupy the same quantum state.” We can state this principle more generally as “No two identical fermions can occupy the same quantum state.”

Bosons are defined as particles that do not obey the Pauli exclusion principle. In a quantum system consisting of many identical bosons, an unlimited number of the bosons can exist in the same quantum state. One example of a boson is the alpha particle, which is the nucleus of a helium atom consisting of two protons and two neutrons. Another example is the deuteron, which is a nucleus with one proton and one neutron. Note that the term “particle” is not limited to “elementary” particles such as electrons, but includes complex bound collections of particles. For instance, a neutral atom of rubidium 87, consisting of 37 protons, 37 electrons, and 50 neutrons, can be regarded as a “particle” that happens to be a boson. The deuterium atom, which is an isotope of hydrogen consisting of an electron (a fermion) bound to a deuteron (a boson), is a fermion.

How does one tell if a particle is a boson or a fermion? One way, in principle, is to test whether the particle obeys the Pauli exclusion principle; that is, put many of the particles in a box and see if more than one can occupy a given quantum state. If only one particle can occupy each state, the particle is a fermion; if more than one can, then it must be a boson. Fortunately, there is an easier way to distinguish between fermions and bosons — the *spin* of the particle determines its type. We saw in Chapter 9 that electrons have an intrinsic angular momentum called the spin. All particles, no matter how complex, have a spin, that is, an internal total angular momentum. Like energy, the internal angular momentum of a particle or of a system of particles is always quantized; the maximum *z* component of the angular momentum can always be written as $s\hbar$, where the number *s*, which is called the spin, is either an integer ($0, 1, 2, 3, \dots$) or a half integer ($\frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, etc.). Experimentally, it is found that if the total spin is an integer, the particle is a boson; if the spin is half-integer, the particle is a fermion. (This correspondence between spin and the Pauli principle can be derived from relativistic quantum theory.) Electrons, protons, and neutrons all have spin $\frac{1}{2}$, so they are fermions. Composite particles, such as atoms, can be either fermions or bosons. In general, if an atom contains an even number of fermions, it is a boson. For instance, in the ground state of the rubidium 87 atom, the spins of the 37 protons, 37 electrons, and 50 neutrons add up in such a way that the angular momenta of all ($37 + 37 + 50 = 124$) spin-half fermions cancel and produce a boson with spin zero. If an atom has an odd number of fermions, then its total spin is half-integer and the atom as a whole is a fermion.

If a collection of identical particles is placed in a rigid box or some other quantum potential well, then which quantum states are occupied by the particles



442 Chapter 13 • Solids — Theory

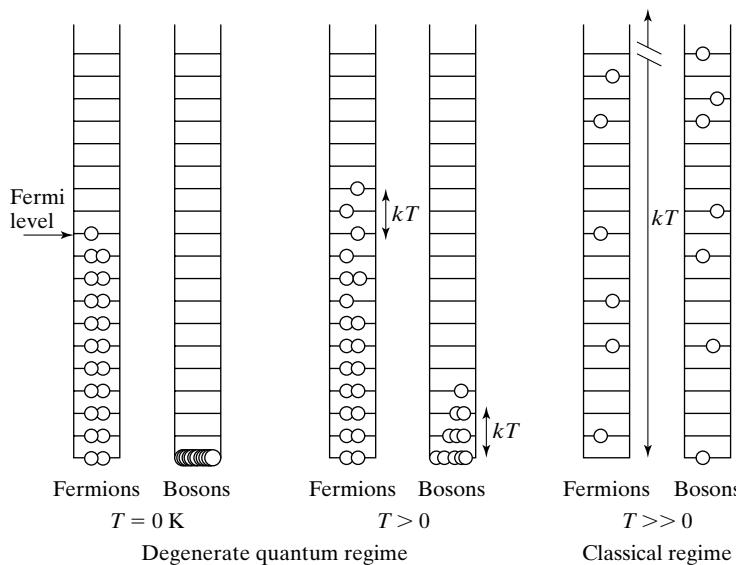


FIGURE 13.36

Fermion and boson gases at various temperatures.

depends both on the temperature and on whether the particles are fermions or bosons. At lower temperatures the particles tend to occupy the available states with the lowest energy, while at high temperatures the particles tend to spread out in energy and occupy states of both low and high energy with nearly equal probability. In Figure 13.36, we compare the behavior of a gas of fermions and a gas of bosons. The figure shows the single-particle energy levels of a quantum system, such as a rigid box or a harmonic oscillator or some other, more complex potential. These levels are called “single-particle states” because they are the solutions of the time-independent Schrödinger equation for any one particle. If we assume that the potential energy function $U(x, y, z)$ in the Schrödinger equation does not involve the spin of the particle, the levels are degenerate. For the case of an electron, each level is at least doubly degenerate and represents two quantum states, one with spin up and one spin down.

We now imagine filling these quantum states with identical particles, either bosons or fermions. We assume the particles are noninteracting, so the presence of other particles does not affect the potential that a given particle experiences, and the solutions of the one-particle Schrödinger equation remain valid, even though there is more than one particle present. At absolute zero ($T = 0 \text{ K}$), the particles will occupy the lowest available states. For the case of spin- $\frac{1}{2}$ fermions, the Pauli exclusion principle requires that each orbital be occupied by no more than 2 fermions, so the states are completely filled up to a maximum level, the Fermi level. However, for bosons, the Pauli principle does not apply; *all* the particles can happily occupy the *same* quantum state, and at $T = 0$, the bosons all crowd into the lowest energy state, the ground state. As we have seen in previous sections, examples of degenerate Fermi gases include conduction electrons in metals and in white dwarf stars, as well as neutrons in neutron stars. Degenerate Fermi gases are common. In contrast, the degenerate boson gas, also called a “Bose gas” or “Bose–Einstein condensate,” is a very exotic state and is found only at extremely low temperatures. As far as we know, degenerate Bose gases do not occur in nature. Some human-made examples of a degenerate Bose gas are described in the next section.

Still looking at Figure 13.36, we now imagine increasing the temperature of our gas slightly. As explained in Chapter 3, at temperature T , a particle has a **thermal energy** of about kT , where k is Boltzmann’s constant. This results in

a *thermal smearing* of the occupancy of the levels. For fermions, some of the states above the Fermi level are now occupied, due to thermal activation of the particles, leaving some of the states below the Fermi level vacant. For bosons, not all the particles are in the ground state; some occupy excited states about kT above the ground state.

Now let us raise the temperature considerably. With the temperature high enough, regardless of whether the particles are fermions or bosons, they are now widely dispersed among the quantum states. The average occupancy of every state is much less than one, and it is unlikely that more than one particle will compete for occupancy of the same state. In this high temperature regime a collection of particles behaves as a classical gas because quantum effects due to multiple occupancy are negligible; in this classical regime a gas of bosons behaves just like a gas of fermions. In contrast, when the temperature is sufficiently low that the occupancy of the states depends on whether the particles are fermions or bosons, we are in the degenerate regime. In this low-temperature regime, the collection of particles is referred to as a *quantum* gas.

The temperature required to reach the classical regime depends on the density of the particles in the system. The gas is in the classical regime if the density is low enough so that particles can “stay out of each other’s way,” that is, if their quantum-mechanical wave packets do not overlap. We can estimate the temperature and density necessary for this classical behavior by the following argument, which has three main ingredients:

1. If the number of particles per volume is n and the average distance between neighboring particles is r , then on average, the volume per particle is r^3 , and

$$n \approx \frac{1}{r^3} \quad (13.27)$$

2. A free particle with kinetic energy E and momentum p has a de Broglie wavelength λ related to E and p by

$$E = \frac{p^2}{2m} = \frac{\hbar^2}{2m\lambda^2} \quad (13.28)$$

3. For a free particle at temperature T , the kinetic energy is given by

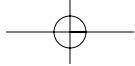
$$E = \frac{1}{2}mv^2 = \frac{3}{2}kT \quad (13.29)$$

where v is the mean thermal speed. This relation was discussed in Section 3.7.

Wave packets have a size of at least the de Broglie wavelength λ , so the condition of nonoverlapping wave packets can be written $r \gg \lambda$; that is, the mean distance between particles is much greater than their wavelength, as seen in Figure 13.37. We can combine points (2) and (3) [equations (13.28) and (13.29)] to obtain an expression for the de Broglie wavelength:

$$\frac{3}{2}kT = \frac{\hbar^2}{2m\lambda^2} \quad \text{or} \quad \lambda = \frac{\hbar}{\sqrt{3mkT}} \quad (13.30)$$

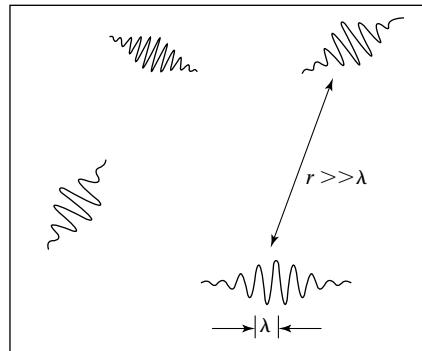
This last expression is called the **thermal de Broglie wavelength**; it is the wavelength of particles at a temperature T . At high temperatures the de Broglie wavelength is small and the particle can be represented by a small, compact wave packet that behaves very much like a classical particle. At very low temperatures, however, the wavelength is large, and the wave packets describing the particles may overlap or even fill the volume of their container.



444 Chapter 13 • Solids — Theory

FIGURE 13.37

Atomic wave packets in a low-density gas. If the wave packets do not overlap, the gas is in the nondegenerate classical regime.



Finally, we combine observation (1) [Eq. (13.27)] with (13.30) to write the classical gas condition $r \gg \lambda$ as

$$\frac{1}{n^{1/3}} \gg \frac{h}{\sqrt{3mkT}} \quad \text{or} \quad n \ll \left(\frac{3mkT}{h^2} \right)^{3/2} \quad (13.31)$$

We have achieved our goal. This last equation is the condition required for a gas to be in the classical regime.

We can define a **quantum concentration** n_q as

$$n_q = \left(\frac{3mkT}{h^2} \right)^{3/2} \quad (13.32)$$

which is the approximate particle concentration at the crossover from classical to quantum regimes. At concentrations higher than this, the wave packets overlap strongly, quantum effects cannot be ignored, and the gas is in the degenerate quantum regime. At concentrations much less than this, the gas is in the classical regime and the ideal gas law applies.

Example 13.6

What are the thermal de Broglie wavelengths of free electrons and neutrons at room temperature?

At room temperature, $T \approx 300$ K. From Eq. (13.30) we have, for electrons

$$\begin{aligned} \lambda &= \frac{h}{\sqrt{3mkT}} = \frac{6.63 \times 10^{-34}}{\sqrt{3 \times (9.11 \times 10^{-31}) \times (1.38 \times 10^{-23}) \times (300)}} \\ &= 6.2 \times 10^{-9} \text{ m} \approx 6 \text{ nm} \end{aligned}$$

For neutrons,

$$\begin{aligned} \lambda &= \frac{h}{\sqrt{3mkT}} = \frac{6.63 \times 10^{-34}}{\sqrt{3 \times (1.67 \times 10^{-27}) \times (1.38 \times 10^{-23}) \times (300)}} \\ &= 1.5 \times 10^{11} \text{ m} \approx 0.015 \text{ nm} \end{aligned}$$

Note that for electrons, the room-temperature wavelength is much greater than the distance between atoms in a solid. Thus, for conduction electrons in a solid, the wave packets strongly overlap, and they must form a degenerate quantum gas.

Example 13.7

Consider the degenerate Fermi gas formed by conduction electrons in a metal. How high the temperature be raised for the gas to enter the classical regime?

The crossover from degenerate Fermi gas to nondegenerate classical gas occurs when the number density (or concentration) n is equal to the quantum concentration n_q given by (13.32). Solving the equation $n = n_q$ for the temperature T , we obtain

$$T = \frac{\hbar^2 n^{2/3}}{3mk} \quad (13.33)$$

Since there is about 1 conduction electron per atom in a metal, the number density of conduction electrons is roughly the same as the density of atoms, and we have $n = 1/a^3$, where a is the lattice constant, the distance between nearest-neighbors atoms in the metal. For all solids, the lattice constant is roughly $a = 0.3$ nm. We now substitute $n = 1/a^3$ into (13.33) and obtain

$$\begin{aligned} T &= \frac{\hbar^2}{3mk a^2} \\ &= \frac{(6.6 \times 10^{-34})^2}{3 \times (9.1 \times 10^{-31}) \times (1.4 \times 10^{-23}) \times (0.3 \times 10^{-9})^2} \approx 130000 \text{ K} \end{aligned} \quad (13.34)$$

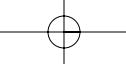
This temperature is far above the melting point of any solid. We conclude that for any solid metal, the temperature is always well below that needed to excite its electrons into the classical regime. Conduction electrons are always in the strongly degenerate quantum regime.

13.12 Bose–Einstein Condensation*

*This material can be omitted without loss of continuity.

In 1924 Einstein predicted that a gas of identical bosons would undergo a new kind of phase transition if cooled to a low enough temperature. Building on earlier work by the Indian physicist S. Bose (after whom bosons are named), Einstein showed that below a certain critical temperature, most of the bosons in a gas will crowd into the ground state. This low-temperature collection of ground-state bosons is called the **Bose–Einstein condensate**.

Before the work of Bose and Einstein, it was already well understood that the particles in a gas of bosons would all enter the ground state when the temperature is very close to absolute zero. Particles in a gas have a thermal energy of about kT , and so we expect that most of the particles will be in the ground state when the temperature is so low that this thermal energy is less than the energy of the first excited state in the system. This temperature turns out to be very very very low — about 10^{-13} K for an ordinary gas (see Problem 13.57). What Einstein showed was that bosons will condense into the ground state at a significantly higher temperature than this. The Bose–Einstein condensation occurs when the particles' wave functions begin to overlap, namely, at the temperature of crossover from the classical to degenerate quantum



446 Chapter 13 • Solids — Theory

Carl Wieman and Eric Cornell
(born 1961, American)



A Bose-Einstein condensate in a gas was first achieved on planet Earth at 10:54 a.m. June 5, 1995, in the laboratory of Wieman(left) and Cornell(right) on the campus of the University of Colorado at Boulder. The original experimental apparatus is now at the Smithsonian Institution. Along with MIT physicist Wolfgang Ketterle, Wieman and Cornell were awarded the 2001 Nobel Prize in physics for their experimental work. A dedicated teacher, Wieman now splits his time between research and developing innovative ways of teaching freshman physics.

regimes described in the previous section. This critical temperature, given approximately by Eq. (13.33), depends on the density of the gas and the mass of the particle, but for a low-pressure gas, it is typically around 10^{-7} K or higher (See Problem 13.58). Einstein's derivation is a bit too subtle to describe here, but what he showed was that the phenomenon of Bose condensation is a consequence of the symmetry of the wave function of indistinguishable particles (described in Section 10.5). He showed that not only *could* bosons occupy the same state, but that below the critical temperature, they actually *prefer to*.

More than 70 years passed before the 1924 prediction of Einstein was verified experimentally. In 1995 two physicists at the University of Colorado, Carl Wieman and Eric Cornell, led a team of researchers who achieved Bose-Einstein condensation (BEC) in a gas of rubidium 87 atoms. In 2001 Cornell and Wieman, along with MIT physicist Wolfgang Ketterle, received the Nobel Prize in physics for their experimental studies of Bose-Einstein condensation.*

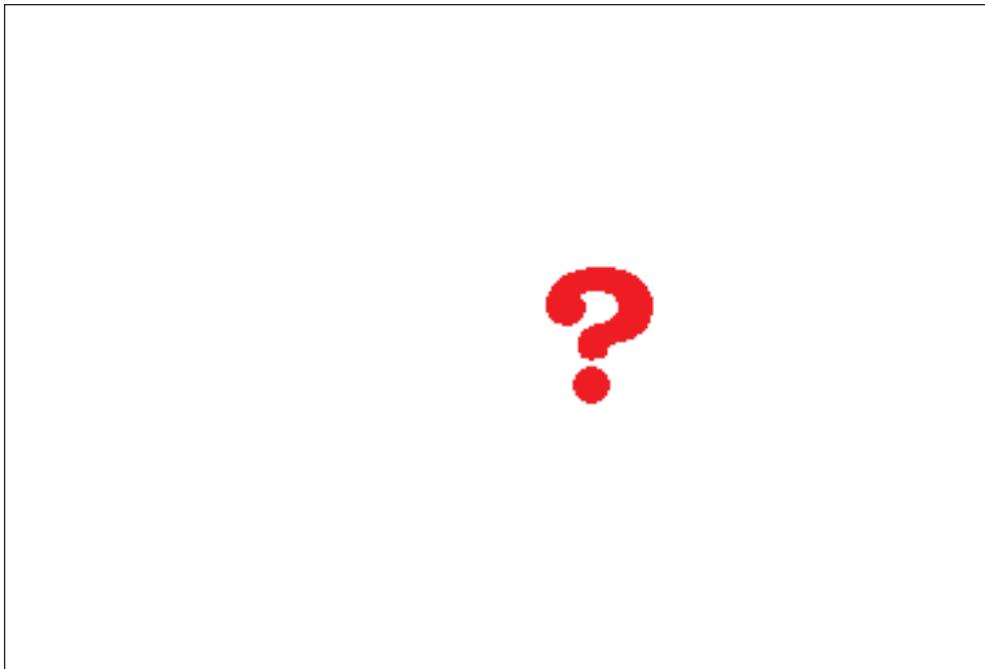
We briefly describe the original experiment of Wieman and Cornell, which is a marvel of ingenuity. The basic problem was how to cool a gas sample down to 10^{-7} K, a temperature that had not been achieved before. They began by suspending the gas sample in a magnetic trap so that the atoms of the gas did not touch the hot walls of the sample container. They then cooled the gas in two stages. In the first stage a technique called laser Doppler cooling was used to cool a sample of about 10^{17} atoms down to 10^{-5} K. In laser Doppler cooling, lasers shine on the sample from six directions (Figure 13.38). The frequency of the laser light is tuned to be near an electronic transition of the atom, so that when the atom is moving toward the laser light, the Doppler-shifted frequency of the light is just right to cause an inelastic collision that slows the atom. Because of the Doppler shift, the atom interacts only with the laser beam it is heading into. In this way, the atom is slowed by one of the six beams, no matter which way it travels, and it gradually cools by transferring its energy and momentum to re-emitted light.

FIGURE 13.38

Schematic of the experimental apparatus of Cornell and Wieman. A 1-inch diameter glass chamber contains a vapor of several thousand rubidium atoms. The vapor is suspended in the center of the chamber by magneto-optical forces; it is illuminated from six sides by laser beams that cool the sample by laser Doppler cooling. The white arrows indicate current directions in surrounding magnets.



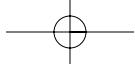
*Another example of experimentalists winning a Nobel Prize for verifying a theoretical prediction due to Einstein. In 1926 French experimentalist Jean Perrin won the prize for verifying Einstein's theory of Brownian motion. In 1923 American experimentalist Robert Millikan won the prize for verifying Einstein's theory of the photoelectric effect. In 1993 American radio astronomers Russell Hulse and Joseph Taylor won the prize for their studies of gravitational radiation from a radio pulsar, which verified Einstein's theory of general relativity. Einstein's theory of special relativity is at the foundation of the field of high-energy physics, which has produced more than two dozen Nobel prizes. Einstein won the prize once, in 1921, for the photoelectric effect.

**FIGURE 13.39**

The Bose–Einstein condensation in a gas of rubidium atoms. These plots indicate atom density (vertically) versus position (horizontally) in the vapor. The field of view of each image is about $200 \times 270 \mu\text{m}^2$. The atom density was measured by shining a laser through the sample and imaging the scattered light. The act of measuring the atom density with this laser flash destroys the sample, so the sample must be re-formed and recooled after each “snapshot.” The sharp peak in atom density below about 200 nK indicates the Bose–Einstein condensation.

In the second stage of cooling the temperature is lowered by controlled evaporation. The magnetic container is adjusted so that the fastest moving (hottest) atoms have just enough energy to escape the sample, leaving behind cooler-than-average atoms. When the original sample of 10^{17} atoms had evaporated down to 2000 atoms, the temperature had dropped to 10^{-7} K and the Bose–Einstein transition occurred. (See Figure 13.39 and Problem 13.58). Virtually all the rubidium atoms dropped into the ground state, and for all practical purposes, the sample was at absolute zero. This experiment was especially elegant because all the equipment was inexpensive and small enough to fit on a tabletop. Cheap diode lasers and hand-wound magnetic coils were used, rather than the large, expensive liquid helium cryostats and superconducting magnets found in most low-temperature laboratories.

The work of Wieman, Cornell, and Ketterle provided the first experimental verification of Bose–Einstein condensation in a gas, that is, a collection of *weakly interacting* particles. However, the basic phenomenon had already been observed in systems of *strongly-interacting* particles, the two clearest examples being superconductivity and superfluidity. **Superconductivity**, which is described in more detail in the next chapter, is a phenomenon that occurs in many metals. When cooled to a few degrees above absolute zero, superconducting metals loose all electrical resistance and become *perfect* electrical conductors. The conduction electrons in these metals interact with the lattice



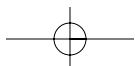
448 Chapter 13 • Solids — Theory

vibrations in a rather complex way and combine to form bound pairs of electrons that have total spin zero and are thus bosons. It is these electron pairs that undergo a Bose–Einstein-like transition at a few degrees Kelvin and cause superconductivity. **Superfluidity** is a phenomenon that occurs in liquid helium. The most common isotope of helium, ${}^4\text{He}$, is a boson. Helium forms a liquid at temperatures below 4.2 K; below 2.2 K, liquid ${}^4\text{He}$ undergoes a Bose–Einstein transition to form a so-called superfluid, which can flow through a pipe with *zero* friction. (See Problem 13.59.) A rare isotope of helium, ${}^3\text{He}$, a fermion, pairs in a manner similar to superconductivity in metals and forms a superfluid below 0.0027 K (2.7 mK). The amazing properties of superconductors and superfluids are a consequence of a large fraction of the particles being in the quantum ground state.

Although superconductors and superfluids are both fairly well understood, a detailed theoretical description has proved enormously difficult and elusive because of complications arising from the strong interactions between particles. The original theory of Bose–Einstein condensation provides a very clear theoretical picture, but really applies only to the simple case of a gas of noninteracting particles. Now that BEC has finally been achieved in a gas, a clean comparison between theory and experiment is possible, and rapid progress is being made in this revitalized field of atomic physics.

CHECKLIST FOR CHAPTER 13

CONCEPT	DETAILS
Types of bonding in solids	Covalent, metallic, ionic, dipole–dipole (Sec. 13.2)
Crystals and noncrystals	Long-range order in crystals, only short-range order in amorphous materials (Sec. 13.3)
Types of crystal structures	Simple cubic, FCC, BCC, diamond, etc. (Sec. 13.3)
Metals, insulators, and semiconductors	Metals have a partly full band of electronic states. Insulators and semiconductors have a full band separated from an empty band by an energy gap. (Sects. 13.4, 13.5)
Drude model of conductivity	$\sigma = ne^2\tau/m$ (13.11)
The drift velocity	Average velocity of electrons in a current-carrying wire — very slow, of order 1 mm/s. (Sec. 13.6)
Electronic collisions in metals	Electron-impurity and electron-phonon scattering (Sec. 13.7)
Ohm's law	$R = V/I = \text{constant}$, $\rho = E/j = \text{constant}$ (Sec. 13.8)
The Fermi energy	Energy of conduction electrons in highest filled energy states. $E_F \approx \hbar^2/(8mr^2) \approx \text{a few eV}$ (13.16)
The Fermi speed	Very high intrinsic speed of conduction electrons due to Pauli exclusion principle, $v_F \approx 10^6$ m/s. (Sec. 13.8)
Degeneracy pressure*	Pressure of degenerate electron gas, due to Pauli exclusion principle, $p \approx (1/3)nE_F$ (13.25)
White dwarfs, neutron stars, black holes*	Stellar remnants: gravity balanced by electron degeneracy pressure, neutron degeneracy pressure, or nothing (Sec. 13.10)
Classical versus quantum gases*	When wave packets overlap, quantum effects are important. Crossover at density $n_q = (3mkT/\hbar^2)^{3/2}$. (13.32)
Bose–Einstein condensation*	Bosons violate Pauli exclusion principle, with a vengeance! (Sec. 13.12)



PROBLEMS FOR CHAPTER 13

SECTION 13.2 (Bonding of Solids)

- 13.1** • From the data in Table 13.1, find the cohesive energies of diamond and germanium in J/mol and kcal/mol.
- 13.2** • From the data in Table 13.1, find the cohesive energies of NaCl and LiF in J/mol and kcal/mol. (Be careful! How many atoms are contained in a mole?)
- 13.3** • The force between two electric dipoles (each of zero net charge) results from the imperfect cancellation of the various attractive and repulsive forces among the four charges involved. **(a)** Prove that the force between the two dipoles in Fig. 13.40(a) is repulsive. **(b)** What is it for the dipoles in Fig. 13.40(b)?

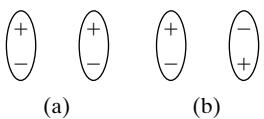


FIGURE 13.40

(Problem 13.3).

- 13.4** •• **(a)** Prove that the force between two dipoles with two like charges closest, as in Fig. 13.41(a), is repulsive. **(b)** What is the nature of the force if unlike charges are closest as in Fig. 13.41(b)?

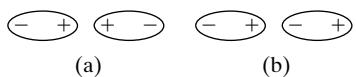


FIGURE 13.41

(Problem 13.4).

- 13.5** •• The van der Waals force between two atoms arises because the fluctuating dipole moment of atom 1 induces a dipole moment in atom 2, and this induced moment is always oriented so as to experience an attractive force from atom 1. This point is illustrated in Fig. 13.5, for one orientation of the dipoles. **(a)** Sketch the electric field lines of dipole 1 in Fig. 13.5, and verify that the induced dipole 2 is oriented as shown. **(b)** Choose axes in the usual way (x across and y up the page) and put the origin at dipole 1. If atom 2 is moved onto the y axis, what is the orientation of its induced dipole moment, and what is the direction of the force on it? **(c)** Repeat for the case that atom 2 is put halfway between the x and y axes.

SECTION 13.3 (Crystals and Noncrystals)

- 13.6** • One can think of a simple cubic lattice [Fig. 13.8(b)] as a stack of many unit cubes [Fig. 13.8(c)]. However, in counting atoms there is a danger of miscounting, since adjacent cubes share several atoms. **(a)** Consider any one atom in the lattice. How many cubes share this atom? **(b)** Consider any one cube. How many atoms does the cube have a share of? **(c)** Prove that in the lattice as a whole, there is one atom per unit cube.

- 13.7** • Consider the Na^+ ion at the center of the cube shown in Fig. 13.9(b). **(a)** In terms of the nearest-neighbor distance r_0 , find the distance of any one of the next-nearest-neighbor ions. **(b)** How many of these ions are there? **(c)** Are they Na^+ or Cl^- ?

- 13.8** • Consider the Na^+ ion at the center of the cube shown in Fig. 13.9(b). **(a)** In terms of the nearest-neighbor distance r_0 , find the distance of any one of the third-nearest-neighbor ions. **(b)** How many of these ions are there? **(c)** Are they Na^+ or Cl^- ?

- 13.9** • The density of solid NaCl is 2.16 g/cm^3 . Find the nearest-neighbor distance. [Hint: The number density of ions is 1 ion per r_0^3 .]

- 13.10** • The KCl crystal has the same FCC structure as NaCl, and the nearest-neighbor distance is 0.315 nm. What is the density of KCl? [Hint: The number density of ions is 1 ion per r_0^3 .]

- 13.11** •• In Example 13.1 [Eq. (13.2)] we estimated the potential energy (PE) of any one ion in solid NaCl. **(a)** Make a corresponding estimate of the PE of any one ion in KCl whose crystal structure is the same as NaCl, but whose nearest-neighbor separation is 0.31 nm. **(b)** This estimate ignores the repulsive forces that come into play when the ions get very close. Thus, your answer should be somewhat below the observed value of -7.2 eV . Use this observed value to find the atomic cohesive energy of KCl in eV/atom. (The ionization energy of K is 4.3 eV ; the electron affinity of Cl is 3.6 eV .)

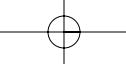
- 13.12** •• Consider the Na^+ ion at the center of the cube in Fig. 13.9(b). **(a)** How many fourth-nearest neighbors does it have? **(b)** Are they Na^+ or Cl^- ? **(c)** Where are they? **(d)** What is their distance?

- 13.13** •• We saw in Example 13.1 that the PE of any one ion in NaCl has the form $U = -\alpha ke^2/r_0$, where $\alpha = 1.75$ is called the Madelung constant. The calculation of this constant in three dimensions is much more trouble than it is worth here, but the one-dimensional Madelung constant is fairly easy to calculate, as follows: Consider a long one-dimensional “crystal” of alternating Na^+ and Cl^- ions spaced a distance r_0 apart. Prove that the PE of any one ion is $U = -\alpha ke^2/r_0$, where $\alpha = 1.39$. [Hint: The series for $\ln(l+z)$ in Appendix B (with $z = 1$) is useful in this calculation.]

- 13.14** •• Extend the reasoning of Problem 13.6 to prove that the number of atoms per unit cell of the BCC lattice is 2.

- 13.15** •• Extend the reasoning of Problem 13.6 to prove that the number of atoms per unit cell of the FCC lattice is 4.

- 13.16** •• Gold has the FCC structure, and its density is 19.3 g/cm^3 . **(a)** Use the result of Problem 13.15 to find the edge length of the unit cube. **(b)** What is the nearest-neighbor separation (center-to-center, as usual)?



450 Chapter 13 • Solids — Theory

13.17 •• Copper has the FCC structure, and the center-to-center distance between nearest-neighbor atoms is 0.255 nm. Use this information and the result of Problem 13.15 to find the density of solid copper.

13.18 •• Iron has the BCC structure, and its density is 7.86 g/cm³. Use the result of Problem 13.14 to find the center-to-center distance between nearest-neighbor iron atoms.

13.19 •• In the CsCl crystal the Cs⁺ ions and Cl⁻ ions are arranged on two identical simple cubic lattices. The Cl⁻ lattice is offset from the Cs⁺ lattice, so that each Cl⁻ is at the exact center of a cube of Cs⁺ ions, and vice versa. Thus a unit cell of CsCl looks just like the BCC cell in Fig. 13.10 (except that the center ion is Cl⁻, whereas those at the corners are Cs⁺, or vice versa). **(a)** The density of CsCl is 3.97 g/cm³. Use the result of Problem 13.14 to find the edge length of the unit cube of CsCl. **(b)** What is the nearest-neighbor distance?

13.20 •• (a) Imagine a large box full of identical rigid spheres packed in a simple cubic lattice and each touching all of its nearest neighbors. What fraction of the total volume is taken up by the spheres? (This fraction is called the *packing fraction*.) Answer the same question for **(b)** FCC and **(c)** BCC packing. **(d)** Which arrangement packs the most spheres per volume?

13.21 •• To estimate the tensile strength of the material in Example 13.2, we had to know the force needed to pull two atoms apart. To get an estimate of this force, consider the semiclassical model of the H₂ molecule in Fig. 12.34 (Problem 12.11). Using the result $s = r_0/\sqrt{3}$ from Problem 12.11 and taking $r_0 = 0.3$ nm as a representative lattice spacing, find the total force of one atom on the other.

13.22 •• Problem 13.21 suggests one way to estimate the force needed to separate two bonded atoms; here is another. The graph in Fig. 12.6 shows the energy of an Na⁺–Cl⁻ pair as a function of their separation. Since the slope of this graph is the force between the ions, the maximum force needed to pull them apart is just the maximum slope of the graph. Use the graph to show that this force is of order 3×10^{-9} N.

13.23 ••• In Example 13.1, Eq. (13.1), we wrote down the first three terms in the electrostatic PE of an ion in the NaCl lattice. What are the next two terms?

13.24 ••• In Example 13.1, Eq. (13.1), we said that the electrostatic PE of any one ion in NaCl is $U_{\text{es}} = -\alpha ke^2/r$, where $\alpha = 1.75$. This is the electrostatic energy of simple point charges, and we should have included a contribution due to the repulsion of the ions at close range. This repulsive contribution can be approximated by a term of the form $U_{\text{rep}} = +\beta/r^n$, where β and n are positive constants. Thus the total PE is

$$U(r) = U_{\text{es}}(r) + U_{\text{rep}}(r) = -\alpha \frac{ke^2}{r} + \frac{\beta}{r^n}$$

(a) At the equilibrium separation $r = r_0$, this energy is a minimum. Use this fact to find the constant β , and hence show that the total PE of an ion at equilibrium is

$$U(r_0) = -\alpha \frac{ke^2}{r_0} \left(1 - \frac{1}{n}\right)$$

(b) This result differs from our rough estimate (13.2) only by the second term $1/n$. For NaCl, $r_0 = 0.281$ nm and $n = 8$. Use these values to find the total PE of an ion in NaCl. Compare with the rough answer of -9.0 eV found in Example 13.1 and the observed value of -7.92 eV.

SECTION 13.4 (Energy Levels of Electrons in Solids; Bands)

13.25 • The density of solid silicon is 2.33 g/cm³. **(a)** Estimate the total number, N , of silicon atoms in a micro-crystal $1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m}$. **(b)** The valence band in silicon is about 1 eV wide. Given that there are $4N$ states in this band, make a rough estimate of the average separation between states in the valence band of this crystal. **(c)** Consider a silicon solar cell that is $1\text{ cm} \times 1\text{ cm} \times 1\text{ mm}$; repeat both calculations for these dimensions.

13.26 • The density of solid germanium is 5.32 g/cm³. **(a)** Estimate the total number, N , of atoms in a micro-crystal $1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m}$. **(b)** The conduction band in germanium is about 2 eV wide. Given that there are $4N$ states in this band, make a rough estimate of the average separation between states in the conduction band of this crystal. **(c)** Consider a germanium crystal that is $1\text{ cm} \times 1\text{ cm} \times 1\text{ cm}$; repeat both calculations for these dimensions.

13.27 • Imagine a crystal containing just one type of atom on a simple cubic lattice. If the lattice spacing is 0.2 nm and the whole crystal is $3\text{ mm} \times 3\text{ mm} \times 3\text{ mm}$, how many states are there in the $2p$ band?

13.28 • Consider a solid containing N atoms and a band that evolved from an atomic level of angular momentum l . How many states does this band contain? Check that your answer gives $2N$ for an s band and $6N$ for a p band.

SECTION 13.5 and 13.6 (Conductors and Insulators — A Qualitative View)

13.29 • Consider a single crystal of an insulator with a band gap of 3.5 eV. What colors of visible light will this crystal transmit?

13.30 • Crystalline sulfur absorbs blue light but no other colors; as a result, it is a pale yellow transparent solid. It is also an electrical insulator. From this information, deduce the width of the band gap of crystalline sulfur and describe the occupancy of the bands above and below the gap.

13.31 • Germanium is a semiconductor with a band gap of 0.7 eV. What colors of visible light does it transmit?

13.32 • Explain why many insulators are transparent to visible light, whereas most semiconductors are opaque to visible light but transparent to infrared.

13.33 • The longest wavelength of electromagnetic radiation that is absorbed in gallium arsenide is 890 nm. What is the band gap of this semiconductor?

- 13.34 ••** Consider a hypothetical element that forms a solid with bands as shown in Fig. 13.42. **(a)** Suppose that the isolated atom has configuration $1s^22s^2$. If its equilibrium separation is $r_0 = a$, is the solid a conductor or an insulator? What if $r_0 = b$? Answer both questions for the case that the atomic configuration is **(b)** $1s^22s^22p^1$ and **(c)** $1s^22s^22p^6$.

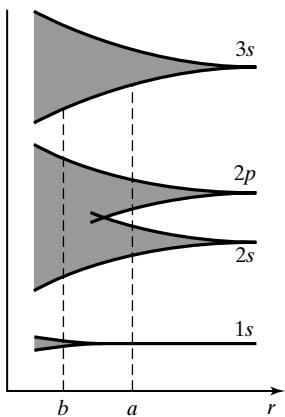


FIGURE 13.42

(Problem 13.34).

SECTION 13.6 (Drude Model of Conductivity)

- 13.35 • (a)** Make a sketch of current I versus voltage V for an ohmic resistor. What is the significance of the slope of your graph? **(b)** When a 120-V lightbulb is on, the temperature of the filament is about 3000°C , and the resistance due to lattice vibrations is approximately 10 times the resistance of the filament at room temperature. Make a sketch of the current I versus voltage V for a tungsten lightbulb filament over a range of voltages from 0 to 120 V. **(c)** Why is a lightbulb most likely to burn out immediately after it is first turned on?

- 13.36 • (a)** Consider a simple cubic lattice of atoms in which the number density of atoms is n_a . Show that the distance a between nearest neighbors is given by $a = n_a^{-1/3}$. **(b)** In a metal we can define a distance r_s as the radius of a sphere whose volume is the volume per conduction electron. Show that r_s is related to the conduction electron density n by the formula $r_s = [3/(4\pi n)]^{1/3}$. **(c)** What is the relation between r_s and a for a monovalent metal with a simple cubic lattice?

- 13.37 •** Compute the number density n (number per volume) of conduction electrons in silver, assuming one conduction electron per atom. The mass density of silver is 10.5 grams/cm³.

- 13.38 ••** Consider a metal in which the distance between nearest-neighbor atoms is $a = 0.3$ nm. **(a)** Assuming a simple cubic lattice structure and one conduction electron per atom, what is the conduction electron density n (number per volume)? **(b)** Assume that a 1-mm-diameter wire made of this metal carries a current of 50 A (a rather large current). What is the current density j ? **(c)** What is the drift velocity in this current-carrying wire?

- 13.39 ••** In Section 13.8 we show that the mean speed of conduction electrons in a metal is roughly $0.003c$ (this is approximately the Fermi speed). Suppose that the velocity of a particular conduction electron is known to within 1%, so that the uncertainty in the speed is approximately $\Delta v = 3 \times 10^{-5} c$. Use the Heisenberg uncertainty principle to show that the resulting uncertainty in the position of the electron, though greater than the distance between nearest-neighbor atoms, is still microscopic. (Since this calculation is only approximate, you can assume 1-D motion.) This calculation is a partial justification for the use of the semiclassical approximation of the Drude model, in which we assume that quantum-mechanical electrons have a well-defined position and momentum.

- 13.40 ••** Consider a cylindrical resistor of length L , cross-sectional area A , resistivity ρ , and resistance R . The resistor has a voltage difference V between its ends and carries a current I . Show that Ohm's law $V = IR$ is equivalent to $E = j\rho$, with resistance R related to the resistivity ρ by the $R = \rho L/A$.

- 13.41 •••** One can define the collision time τ in the following way: The probability P that an electron suffers a collision during an infinitesimal time interval dt is

$$P(\text{collision in time } dt) = dt/\tau \quad (13.35)$$

Note that the probability that the electron suffers *no* collision during a time interval dt is then given by

$$P(\text{no collision in } dt) = 1 - dt/\tau \quad (13.36)$$

Consider now the probability $P'(t)$ that a given electron undergoes no collision during a finite time interval t . **(a)** Show that P' obeys the equation

$$\frac{dP'}{P'} = -\frac{dt}{\tau} \quad (13.37)$$

[Hint: First, argue that $P'(t+dt) = P'(t)(1 - dt/\tau)$.] **(b)** Using (13.37), solve for $P'(t)$ and sketch this function. **(c)** Suppose that a collision occurs at time $t = 0$. Argue that the probability $p''(t) dt$ that the *next* collision occurs between time t and $t + dt$ is given by

$$p''(t) dt = e^{-t/\tau} \frac{dt}{\tau} \quad (13.38)$$

(d) Use (13.38), to compute the average time, $\langle t \rangle$, between collisions. **(e)** Compute the rms average time $\sqrt{\langle t^2 \rangle}$ between collisions.

SECTION 13.7 and 13.8 (Electron Collisions in Metals and The Fermi Speed)

- 13.42 •** Figure 13.29 shows the dependence of resistivity on temperature for a pure metal and a semiconductor. Sketch a new graph, showing the dependence of conductivity on temperature for a metal and a semiconductor. Be sure to label your curves clearly.

452 Chapter 13 • Solids — Theory

13.43 •• At high temperatures ($T \gtrsim 500$ K) the resistivity of tungsten is approximately proportional to the absolute temperature ($\rho \propto T$). A hot tungsten lightbulb filament maintains a stable temperature because the electrical power into the bulb ($P = I^2R = V^2/R$) is exactly balanced by the power out due to cooling by radiation and conduction. At high temperatures the cooling is dominated by radiation, and the power radiated is given by the formula, $P = A \varepsilon \sigma T^4$, where A is the surface area of the body and both ε and σ are constants, called the emissivity and the Stephan–Boltzmann constant,* respectively. (a) Show that the steady-state temperature of a hot lightbulb filament with voltage V is proportional to $V^{2/5}$. (b) Show that the temperature of a hot filament carrying a current I is proportional to $I^{2/3}$.

13.44 •• Estimate the conductivity σ of an alloy or very dirty metal. Assume a simple cubic lattice, one conduction electron per atom, and a mean free path of about 2 lattice constants (twice the distance between atoms). Take the Fermi speed to be $0.003c$, typical of metals. How does this compare to the conductivity of copper at room temperature?

13.45 •• At room temperature the mean free path in a clean metal sample is about 150 nm. The triple-R of the metal [$R(300\text{ K})/R(4\text{ K})$] is measured to be 100. What is the mean free path in the metal at $T = 4\text{ K}$?

13.46 •• Compute the Fermi energy for the case of a divalent metal (one with two conduction electrons per atom). Is the Fermi energy for a divalent metal larger or smaller than for a monovalent metal (all other things being equal)? [Hint: For a simple cubic lattice, the number density n is related to the nearest-neighbor separation r by $n = 1/r^3$.]

13.47 •• Estimate the mean free path in a metal with a resistivity of $5\ \mu\Omega \cdot \text{cm}$ (typical of a good metal at room temperature). Assume a typical metal: one conduction electron per atom, a lattice constant (nearest-neighbor atomic separation) of 0.3 nm, and a Fermi speed of 10^6 m/s . Give the answer in both nm and lattice constants.

13.48 •• (a) Using Drude's formula $\sigma = ne^2\tau/m$ and the known values of the conductivity of metals, make an order-of-magnitude estimate of the collision time τ . **(b)** Compute the *thermal* speed of an electron at room temperature. Drude assumed, incorrectly, that the mean speed of conduction electrons is given by their thermal speed. He also assumed, incorrectly, that it is the ions in a metal that scatter conduction electrons, which implies a mean free path of a few or several lattice constants (depending on the scattering cross section of the ions). **(c)** Use the thermal speed and the collision time to compute the mean free path.

Your answer should be a distance equal to several lattice constants, consistent with Drude's (incorrect) assumption of atomic scattering. This calculation shows that, although Drude made two incorrect assumptions, his model was self-consistent.

SECTION 13.9 (Degeneracy Pressure*)

13.49 • A particle in a 1-D rigid box of length L (an infinite square well) is in its ground state. The box is compressed to length $L/2$. **(a)** By what factor does the ground-state energy increase? **(b)** By what factor does the first excited-state energy increase? **(c)** Repeat the calculations in (a) and (b) for the case of a particle in a 3-D box of volume $V = L^3$.

13.50 •• Equation (13.16) is an approximate expression for the Fermi energy in terms of the mean distance r between nearest-neighbor electrons (which is the same as the distance between nearest-neighbor atoms in a monovalent metal). **(a)** Rewrite this expression to give the Fermi energy as a function of the number density of electrons n (number per volume). **(b)** Consider a metal with a Fermi energy of 4 eV. What is the Fermi energy if this metal is compressed to twice its normal density? (As the next problem shows, such compression requires enormous pressure.) [Hint: You should be able to compute the answer without looking up Planck's constant h or the mass of the electron.]

13.51 •• (a) An ideal gas at standard temperature and pressure (STP: $T = 273\text{ K}$, $p = 1\text{ atm} = 1.01 \times 10^5\text{ Pa}$) is compressed to half of its original volume while the temperature is held constant. What is the pressure of the compressed gas? Give the answer in atmospheres. **(b)** Why is the degeneracy pressure unimportant in an ordinary gas (such as air) at STP? **(c)** A metal with a Fermi energy of 4 eV and an electron degeneracy pressure of 10^6 atm is compressed to half of its initial volume. What is the new value of the degeneracy pressure? [Hint: Use the results of the previous problem.]

SECTION 13.10 (White Dwarfs, Neutron Stars, and Black Holes*)

13.52 • A neutron star consists of neutrons packed so closely together that they touch, just as in the interior of an atomic nucleus. Estimate the density (mass per volume) of a neutron star, given that the radius of a neutron is about 1 fm (10^{-15} m). Compare your answer with the density of water.

13.53 •• (a) Consider a star with a mass 3 times that of the sun, a rotational period of 30 days, and an average density equal to that of water (which is a typical density for a star). If the star collapses into a neutron star with a diameter of 10 km, what is the new rotational period? Assume that the original star and the final neutron star both have uniform densities. [Hint: Use conservation of angular momentum.] This calculation produces a higher rotational rate than is actually found in neutron stars — the fastest known pulsar has a period of 1.6 ms. Our calculation is flawed because stars do not have a uniform density. Instead, most of

*Beware: The Greek letter σ (sigma) is used to denote both the conductivity and the Stephan–Boltzmann constant. Don't confuse them.

the mass is concentrated near the center. Also during the final collapse to a neutron star, the outer layers of the star are blown off, carrying away much of the angular momentum.) **(b)** For a neutron star of 10-km diameter, what is the shortest possible period of rotation compatible with the universal speed limit (the speed of light)?

- 13.54 •• (a)** Use nonrelativistic classical mechanics to derive an expression for the radius of the event horizon surrounding a black hole of mass M . (In this case a nonrelativistic calculation happens to produce exactly the right answer.) **(b)** What is the event horizon radius of a 10-solar-mass black hole? **(c)** What is the acceleration of gravity at the event horizon of a 10-solar-mass black hole?

- 13.55 ••** Derive an expression for the acceleration of gravity at the event horizon of a black hole of mass M . Curiously, the acceleration of gravity at the event horizon *decreases* as the mass of the black hole increases. How massive is a black hole that has an acceleration of g (9.8 m/s^2) at its event horizon? Give your answer in solar masses.

- 13.56 ••• Estimation of the radius of a neutron star.** The following nonrelativistic calculation gives approximately correct values, but a rigorous calculation requires a relativistic treatment. **(a)** Estimate the gravitational pressure at the center of a star of mass M and radius R , assuming a uniform mass density. A fairly accurate estimate can be made from the following rough calculation: Consider the star to be made up of two hemispheres each of mass $M/2$. The centers of mass of the two hemispheres are roughly a distance R apart. The gravitational force between the hemispheres is then approximately $F = G(M/2)^2/R^2$. This force is spread out over the area of contact A between the hemispheres, and the pressure at the center is roughly $p = F/A$. **(b)** Starting with equations (13.16) and (13.25), derive an expression for the neutron degeneracy pressure in terms of R and M . [Hint: The mass density ρ of the star, the number density n of neutrons,

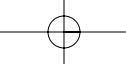
and the neutron mass m are related by $\rho = mn$.] **(c)** The neutron star has a stable size because the inward gravitational pressure is equal to the outward degeneracy pressure. Derive an expression for the radius of a neutron star of mass M . If your expression is correct, it should show that the radius R and the mass M are related by $R \propto M^{-1/3}$. A larger mass produces a smaller radius! **(d)** Calculate the radius of a 2-solar-mass neutron star.

SECTION 13.12 (Bose-Einstein Condensation*)

- 13.57 •• Separation of energy levels in a gas. (a)** Review Sections 7.4 and 8.3, and then compute the separation between the ground-state energy and the first excited-state energy for a particle in a rigid cubical box of edge length L . **(b)** Consider a gas of noninteracting bosons with mass m equal to that of helium atoms, contained in a cubical box, with edge length $L = 1 \text{ cm}$. If no Bose-Einstein condensation were to occur, how low would the temperature have to be for most of the atoms to be in the ground state?

- 13.58 •• Bose-Einstein condensation in rubidium vapor.** Use Eq. (13.33) to estimate the temperature at which a gas of 2000 rubidium 87 atoms confined to a volume of 10^{-15} m^3 will undergo a Bose-Einstein transition. These were the conditions in the original experiment of Cornell and Wieman.

- 13.59 •• Critical temperature for superfluid ${}^4\text{He}$.** Liquid ${}^4\text{He}$ undergoes a transition to a superfluid at a critical temperature T_c , which should be given roughly by (13.33). [Eq. (13.33) was derived for the case of a gas of weakly interacting particles and is not expected to be very accurate for a system of strongly interacting particles such as liquid helium.] Use (13.33) to get an order-of-magnitude estimate of the temperature at which liquid helium undergoes a Bose-Einstein transition and forms a superfluid. The mass density of liquid helium is 0.15 grams/cm³. (The observed value for the critical temperature of liquid helium is $T_c = 2.17 \text{ K}$.)



Chapter 14

Solids—Applications

- 14.1 Overview
- 14.2 Semiconductors
- 14.3 The *pn* Junction Diode
- 14.4 The Transistor
- 14.5 Further Semiconductor Applications*
- 14.6 Integrated Circuits*
- 14.7 The Scanning Tunneling Microscope*
- 14.8 Superconductivity*
- 14.9 The Digital Information Age*

Problems for Chapter 14

*These sections can be skipped without loss of continuity.

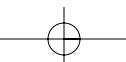
"There has recently been a great deal of thought spent on electronic brains or computing machines. . . . For applications of this sort, there are difficulties in applying vacuum tubes because of their size and the heat which they produce. It seems to me that in these robot brains, the transistor is the ideal nerve cell."

— William Shockley, 1949

14.1 Overview

In this chapter we look at some applications of the solid-state theory described in Chapter 13. Applied solid-state physics is a vast and vitally important subject. The technology arising from solid-state physics — transistors, computers, and digital communications — is changing civilization. We hope that this introduction, though necessarily brief and incomplete, will give you some appreciation of the amazing little slivers of silicon lurking in devices all around you — your watch, car, phone, television, and PC. More and more, these devices are managing our lives.

In Section 14.2 we describe the physics of semiconductors and how the electronic properties of semiconductors can be altered in a controlled way with tiny amounts of impurities. In Section 14.3 we give a semiquantitative view of a simple but fundamentally important solid-state device, the *pn* junction diode. In Section 14.4 we describe the transistor, which has been called the most important invention of the twentieth century. We look at two different transistor types: the bipolar junction transistor, which is used primarily as a discrete component amplifier, and the MOSFET (metal oxide semiconductor field-effect transistor), which is used in integrated circuits as an elementary on/off switch. In Sections 14.5 and 14.6 we examine a few other important applications of semiconductor technology: LEDs, solar cells, CCD cameras, and the crowning glory of the transistor technology — the integrated circuit. In



Section 14.7 we describe the scanning tunneling microscope (STM), invented in 1983, which is now a standard analytical tool in solid-state laboratories around the world. In 14.8 we give an overview of the phenomenon of superconductivity, in which the electrical resistance of some solids drops to zero when the solid is cooled below a certain critical temperature. We conclude this chapter with a look at information theory — how best to store and transmit digital data — and how this has made the digital information age possible.

14.2 Semiconductors

We saw in Section 13.5 that a metal has a partially full conduction band, while an electrical insulator has a large band gap, typically 3 to 5 eV in width, separating a full valence from an empty conduction band. As shown in Fig. 14.1, a **semiconductor** is simply an insulator with a relatively small band gap, around 1 eV or less. In the periodic table semiconductors lie in the vicinity of the group IV elements, between metals on the left and insulators on the right. Table 14.1 lists the band gaps of some semiconductors and insulators. Because the band gap is so narrow in semiconductors, electrons can be thermally excited from the valence band into the conduction band, where they can carry a current.

To understand the thermal activation of electrons in semiconductors, we will have to use a result from statistical mechanics, which is covered in Chapter 15. According to statistical mechanics, for a quantum system with a fixed number of particles in thermal equilibrium at temperature T , the probability $P(i)$ that the system is in a particular quantum state i , with energy E_i , is given by (Eq. 15.2),

$$P(i) \propto e^{-E_i/kT} \quad (14.1)$$

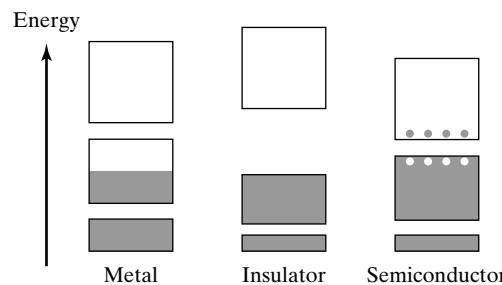


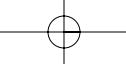
FIGURE 14.1

Band structure of a metal, an insulator, and a semiconductor. The shaded regions indicate occupied electronic states. In insulators a large band gap ($E_g \approx 5$ eV) separates a full valence band from an empty conduction band. In semiconductors the band gap is small enough ($E_g \approx 1$ eV) that electrons can be thermally excited from the valence band to the conduction band.

TABLE 14.1

The band gap (E_g) and dielectric constant for selected semiconductors and insulators. The band gap depends somewhat on temperature; it decreases with increasing temperature. TiO_2 is one of a few insulators with an anomalously large dielectric constant.

	Material	E_g (eV) at T = 300 K	Dielectric Constant
Semi-conductors	Ge	0.66	15.8
	Si	1.17	11.7
	GaAs	1.43	13.1
Insulators	TiO_2	3.03	110
	Diamond	5.4	5.5



456 Chapter 14 • Solids—Applications

where k is Boltzmann's constant and T is the absolute temperature in kelvins. From this relation, it is possible to derive the fact that the concentration n (number per volume) of conduction electrons thermally excited into the conduction band is proportional to an exponential factor* involving the width of the band gap E_g ,

$$n \propto e^{-E_g/(2kT)} \quad (14.2)$$

Recall from Section 3.7 that at room temperature the thermal energy is roughly $kT \approx 1/40$ eV = 0.025 eV. Because of the exponential dependence on the ratio $E_g/(2kT)$ in (14.2), a band gap of 1 eV in a semiconductor at room temperature results in a significant concentration of conduction electrons, while for an insulator with a band gap of 3 eV or more, the concentration of thermally activated carriers is essentially zero.

Example 14.1

At room temperature germanium has a band gap of $E_g = 0.67$ eV and a conduction electron concentration of about $n = 2 \times 10^{13}$ cm⁻³. (Note how small this is compared to range of concentrations in metals, 10^{22} to 10^{23} cm⁻³.) Estimate the conduction electron concentration in diamond, which has a band gap of 5.4 eV.

Using equation (14.2), we can set up an approximate ratio of conduction electron concentrations in two different materials, which we label 1 and 2.

$$\frac{n_2}{n_1} \approx \frac{e^{-E_2/2kT}}{e^{-E_1/2kT}} = e^{-(E_2 - E_1)/2kT} \quad (14.3)$$

This relation is only approximate because the proportionality constant in (14.2) varies from material to material by a factor of 10 or so; however, the exponential factor usually dominates the behavior, as this example will show.

Comparing germanium and diamond, the band gap difference is 5.4 eV – 0.7 eV = 4.7 eV. With $kT_{\text{room}} = 0.025$ eV, equation (14.3) becomes

$$n_{\text{diamond}} \approx n_{\text{Ge}} e^{-(4.7)/(2 \times 0.025)} = (2 \times 10^{13} \text{ cm}^{-3}) e^{-94} \approx 10^{-28} \text{ cm}^{-3}$$

This is such an infinitesimal concentration (much less than 1 electron per cubic kilometer of diamond!) that the conduction electron density in pure diamond can be taken to be zero. Actual diamond samples have a carrier concentration much greater than this (still tiny), but this is due to unavoidable impurities, not thermal activation across the band gap.

A pure semiconductor, such as pure silicon or pure GaAs, is often called an **intrinsic semiconductor**. In contrast, an **impurity semiconductor** is one that

*The factor of 2 in the ratio $E_g/2kT$ arises from the fact that the Fermi energy is in the middle of the band gap. Deriving this would take us too far afield for the present discussion. Afficionados may note that the relation (14.1) applies to the semiconductor as a whole, since it is a multiparticle quantum system in which the number of particles is constant and which can occupy any of its multiparticle quantum states without restriction; however, it does not apply to individual electrons in the semiconductor since electrons, being fermions obeying the Pauli principle, can occupy only single-electron states that are not already occupied by other electrons.

has a small concentration of impurities, added in a controlled way in a process called **doping**. In an intrinsic semiconductor, the concentration of charge carriers is due entirely to activation across the band gap. In an impurity semiconductor, however, we will see that the conduction is controlled by the number and type of impurities, and that by combining two or more semiconductors that are differently doped, one can make an amazing variety of electronic devices. Although it is the impurity semiconductors that are commercially important, we need to say a little more about intrinsic semiconductors first.

Electrons and Holes

If an electron in the valence band of an intrinsic semiconductor is thermally excited into the conduction band, it leaves behind an unoccupied state, or **hole**, in the valence band. The removal of the electron from the valence band leaves behind the unbalanced charge of a positive ion, and so the hole has a positive charge associated with it. Like negatively charged electrons in the conduction band, the positively charged holes in the valence band are mobile and can carry electrical current, as indicated in Fig. 14.2. When an electric field is applied to the right, an electron in the conduction band moves to the left, producing a conventional current to the right. In the valence band, an electron just to the right of a hole can move to the left into the unoccupied state, which shifts the hole to the right. This allows the next electron to the right to move, then the next, and so on. A hole thus behaves like a mobile positive charge, which is free to move around in the valence band and carry a current, just as a negative electron is free to move around in the conduction band.* A current flowing to the right in an intrinsic semiconductor is made up of two components: negative electrons moving to the left in the conduction band and an equal number of positive holes moving to the right in the valence band. This notion of positive holes moving in the valence band will play a central role in our discussion of impurity semiconductors.

The behavior of a hole in the valence band is similar to the behavior of a bubble in water. As shown in Fig. 14.3, when an electron falls down into lower energy state, the hole it fills is displaced upward, and so while electrons tend to fall down into lower energy states, holes tend to rise up toward higher energy states. Thus, holes tend to congregate near the top of the valence band while conduction electrons cluster near the bottom of the conduction band.

Valence holes and conduction electrons can be created by photons, as well as by thermal fluctuations. A photon with energy greater than the band gap ($hf > E_g$) can excite an electron from the valence to the conduction band, as in Fig. 14.4(a); we say that the photon has created an electron-hole pair. Conversely, when an electron drops from the conduction band into a hole in the valence band, it can emit a photon, as in Fig. 14.4(b), and we say that the electron and hole have recombined or annihilated. Electron-hole recombination can create a photon only in so-called “direct-gap” semiconductors, such as GaAs and InSb. In direct-gap semiconductors, electrons at the bottom of the conduction band have nearly the same momentum as holes at the top of the

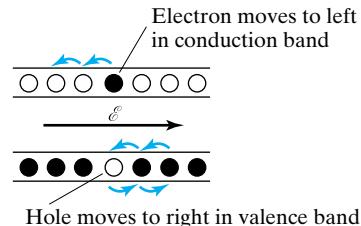


FIGURE 14.2

When a valence electron is excited to the conduction band, it leaves behind an unoccupied valence state, called a hole. An electric field to the right makes conduction electrons move to the left, and holes move to the right as if they were positively charged particles.

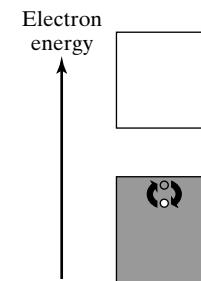


FIGURE 14.3

Electrons tend to fall down into lower energy states. As a result, holes tend to rise to higher energy states. In an electronic energy-level diagram, holes tend to float upward, like bubbles in water.

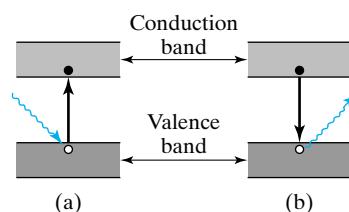
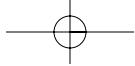


FIGURE 14.4

(a) When a photon excites an electron from the valence band to the conduction band, it creates an electron-hole pair. (b) The conduction electron can recombine with the hole and emit a photon.

*This heuristic discussion of hole conduction is misleading in one respect. When an electric field is applied to the right, electrons in the valence band actually move to the right (as if they were positively charged!) because of a subtle interaction between the electrons and the lattice of positively charged ions. Valence electrons are said to have a “negative effective mass.” The conclusion that holes behave as mobile positive charges remains correct, however.



458 Chapter 14 • Solids—Applications

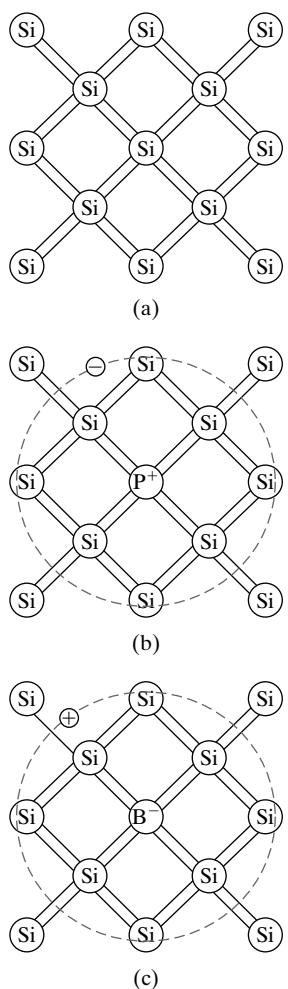


FIGURE 14.5

(a) Schematic view of solid silicon. Silicon has valence 4, and each atom is covalently bonded to its four nearest neighbors. (b) A donor impurity. When a pentavalent impurity, such as phosphorus, substitutes for a silicon atom, four of the five electrons of the impurity complete the covalent bonds. The fifth electron is weakly bound to the P⁺ ion and can be easily excited into the conduction band. (c) An acceptor impurity. A trivalent impurity, such as boron, has only three electrons, so one bond is incomplete. The incomplete bond is a hole, weakly bound to the B⁻ ion. The hole can readily accept an electron from the valence band, creating a mobile hole. (The Bohr radius of the bound electron or hole is actually much larger than shown here.)

valence — a necessary condition for the emission of a photon, as described in Problem 14.17. In silicon and germanium, electron-hole recombination creates phonons (lattice vibrations), which heat the sample. The interaction between photons and electron-hole pairs is the basis of several semiconducting devices including light-emitting diodes (LEDs) and solar cells, as discussed further in Section 14.5.

Impurity Semiconductors

Most commercial applications of semiconductors involve impurity semiconductors, into which small concentrations of carefully chosen impurities have been introduced. The impurities, or *dopants*, can create either conduction electrons or valence holes, depending on the type of dopant. Impurity semiconductors are classified as either **n-type** or **p-type**. In *n*-type semiconductors the impurities contribute electrons to the conduction band, making possible conduction by negative (*n*) carriers; in *p*-type semiconductors the impurities introduce holes into the valence band, allowing conduction by positive (*p*) carriers.

To understand how dopant atoms create electrons or holes, we must examine the crystal structure of the host semiconductor. To be specific, we will look at silicon, which is the semiconductor used in the manufacture of integrated circuits. Silicon is valence 4, and solid silicon forms a diamond crystal structure; every Si atom sits at the center of a tetrahedron and forms four covalent bonds with its four nearest neighbors, as indicated in Fig. 14.5(a). Each covalent bond contains two electrons, shared by neighboring Si atoms. If a valence 5 atom, such as phosphorus, is added to the Si lattice, as in Fig. 14.5(b), it will substitute for one of the Si atoms. Four of the five valence electrons of the P atom will fill the four bonds with the Si neighbors, while the fifth electron will be weakly bound to the P⁺ ion. The binding energy of the extra electron is quite small, about 0.05 eV, and consequently it orbits the P⁺ ion in a wide orbit, with a radius of about 3.0 nm. You might naively expect that the extra electron would form a hydrogen-like orbit about the charge +1e impurity site, with a binding energy around 13.6 eV and a Bohr radius of 0.05 nm. However, the Si lattice is a polarizable medium; the electrons in the covalent bonds are pulled toward the P⁺ ion and partially shield it. The result is that the fifth electron sees an effective charge much less than +1e; for this and other reasons, it is quite weakly bound. (See Problem 14.16 for details.) At room temperature and above, thermal energy ($kT \approx 0.025$ eV) is large enough compared to this binding energy for the weakly bound electron to be excited into an unbound state, leaving an ionized impurity site in the lattice. The unbound electron is a conduction electron, free to move throughout the lattice. Thus, the weakly bound state of the pentavalent impurity has an energy just below the bottom edge of the conduction band, in the band gap, as shown in Fig. 14.6. Pentavalent impurities, such as phosphorus or arsenic, are called **donors** because they donate electrons to the conduction band.

While pentavalent impurities add electrons to the conduction band, trivalent impurities add holes to the valence band. If a valence 3 impurity, such as boron, aluminum or gallium, is added to a silicon lattice, as in Fig. 14.5(c), the three electrons of the impurity complete three of the four covalent bonds surrounding the site. The missing fourth electron constitutes a hole, which is weakly bound to the -1e impurity site. This impurity site forms an electronic state just above the top of the valence band. At room temperature

the bound hole is excited *downward* into the valence band, becoming a mobile hole and leaving behind $-1e$ impurity. (Recall that for holes, higher energy means moving downward among electronic energy levels.) Trivalent impurities are called **acceptors** because they can accept electrons from the valence band and create holes.

Here is another way of understanding why donor states lie just below the conduction band and acceptor states lie just above the valence band. Consider a valence 5 donor impurity, surrounded by four complete covalent bonds, and now ask what happens when we add the fifth electron. Suppose, for just a moment, that instead of a +5 impurity, the site contained an ordinary +4 Si ion. Then the extra electron would go into the lowest available electronic state of a pure Si lattice, namely, a state at the bottom of the conduction band. However, instead of a +4 Si, the site contains a +5 impurity, so the extra electron is a little more strongly bound by the Coulomb attraction of the extra charge. This lowers the energy of the state below the bottom of the conduction band.

We can repeat the argument for a trivalent acceptor site. Consider a valence 3 acceptor impurity, surrounded by *four* completed covalent bonds. An extra electron is required to complete the last bond. We now remove the extra electron, and ask, What is the energy of the resulting hole state? If instead of a +3 impurity, there was a +4 Si at the site, then we would simply have a pure Si lattice with one missing electron; the lowest energy state of such a system has a hole at the top of the valence band. However, the site has a +3 impurity, not a +4 Si, so binding energy of the extra electron is less. The energy of the state is thus a little higher than expected and is raised above the top of the valence band.

Very small concentrations of impurities have a large effect on the carrier concentration in semiconductors. At room temperature and above, each donor impurity adds a conduction electron and each acceptor impurity adds a valence hole. Because the thermally activated carrier concentration in intrinsic semiconductors is very small, even low concentrations of impurities can have a huge effect. For instance, adding just one donor impurity for every million Si atoms increases the room-temperature concentration of mobile charge carriers by a factor of 10^6 , from the intrinsic value of 10^{16} m^{-3} to 10^{22} m^{-3} . Furthermore, the relatively large concentration of conduction electrons in *n*-type semiconductors greatly suppresses the concentration of holes because the electrons drop down into the valence band and fill the holes. Likewise, the large concentration of holes in *p*-type semiconductors greatly suppresses the conduction electron concentration. Thus, impurity doping controls both the concentration and type of charge carriers. The suppression of so-called **minority carriers** — holes in *n*-type and electrons in *p*-type — and the dominance of the **majority carriers** — electrons in *n*-type, holes in *p*-type — play an important role in the behavior of semiconductor devices, as we will see in the following sections.

Because of the extreme sensitivity of semiconductor properties to impurity levels, the manufacture of semiconductor devices must begin with the production of exceedingly pure single crystals of silicon. Impurity levels must start below 1 part in 10^{10} , so that donors or acceptors can then be added in a controlled manner (see Problem 14.12). This level of sample purity was practically unknown prior to the rise of the semiconductor industry, and special purification techniques had to be developed to produce the necessary semiconductor purity.

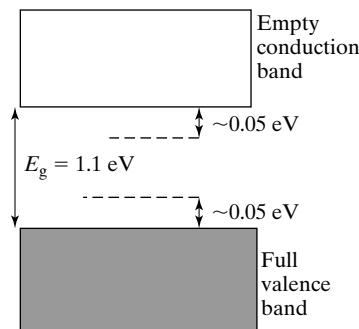
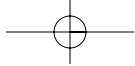


FIGURE 14.6

Impurity states in silicon. Donor states lie just below the bottom of the conduction band, and acceptor states lie just above the top of the valence band. The values shown are for silicon. In this diagram, horizontal position represents position in the silicon lattice, and the short, dark lines represent the electronic states associated with impurity sites scattered throughout the lattice.



460 Chapter 14 • Solids—Applications

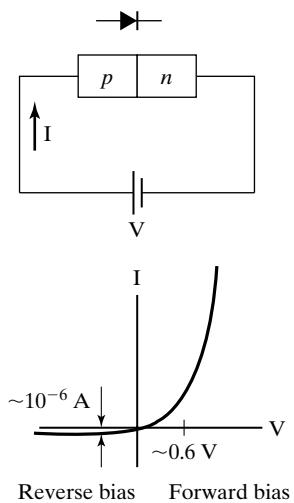


FIGURE 14.7

A diode consists of a single crystal semiconductor that has been doped *p*-type on one half and *n*-type on the other. The symbol at the top is the standard symbol for a diode. The graph shows a typical current-voltage characteristic for a silicon diode.

14.3 The *pn* Junction Diode

The simplest of the semiconductor devices is the diode, which consists of a single-crystal semiconductor that has been doped to make part of it *p*-type and the remainder *n*-type; the interface between the *p*-type and *n*-type is called a ***pn* junction**. As we will see, a diode has the remarkable property that it can easily conduct current in one direction only. A diode is therefore a one-way valve, which can be used, for example, to rectify an AC voltage to DC. The current-voltage characteristic of a typical diode is shown in Fig. 14.7.

To understand the workings of the *pn* junction, let us begin by imagining placing *p*-type and *n*-type pieces of silicon near each other. Before bringing the two types in contact, the energy-level diagrams showing the occupancy of the electronic states at room temperature are as depicted in Fig. 14.8. Shown on the right, the *n*-type silicon has ionized donor states just below the bottom edge of the conduction band, and mobile electrons, donated by those impurities, in the conduction band. On the left the *p*-type silicon has ionized acceptor states just above the valence band and mobile holes in the valence band. Both of these isolated semiconductors are electrically neutral.

Now let us bring the two doped semiconductors in contact and form a *pn* junction. In the *n*-type, near the junction, electrons in the conduction band can lower their energy by diffusing across the junction and then dropping down to fill the hole states in the *p*-type valence band. This charge transfer makes the left side (*p*-type) negatively charged and leaves the right side (*n*-type) positively charged, raising the energy levels on the left relative to those on the right until further charge transfer is no longer favorable. When equilibrium is established, the situation is as depicted in Fig. 14.9. Near the junction there is a **depletion region** in which there are almost no conduction electrons or holes. In this region the ionized donor and acceptor impurity sites form a charged bilayer that creates a very strong local electric field, typically 10^6 V/m (see Problem 14.25).

FIGURE 14.8

Energy-level diagram showing electronic states in *p*-type and *n*-type semiconductors. In *p*-type semiconductors valence holes are created by negatively charged acceptor impurities. In *n*-type, conduction electrons are donated by positively charged donor impurity sites.

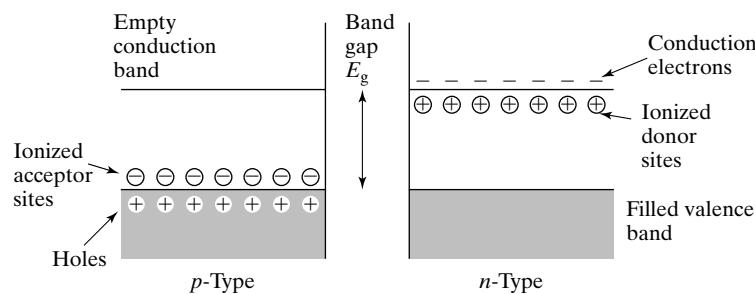
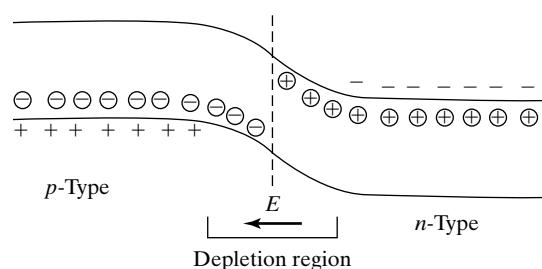


FIGURE 14.9

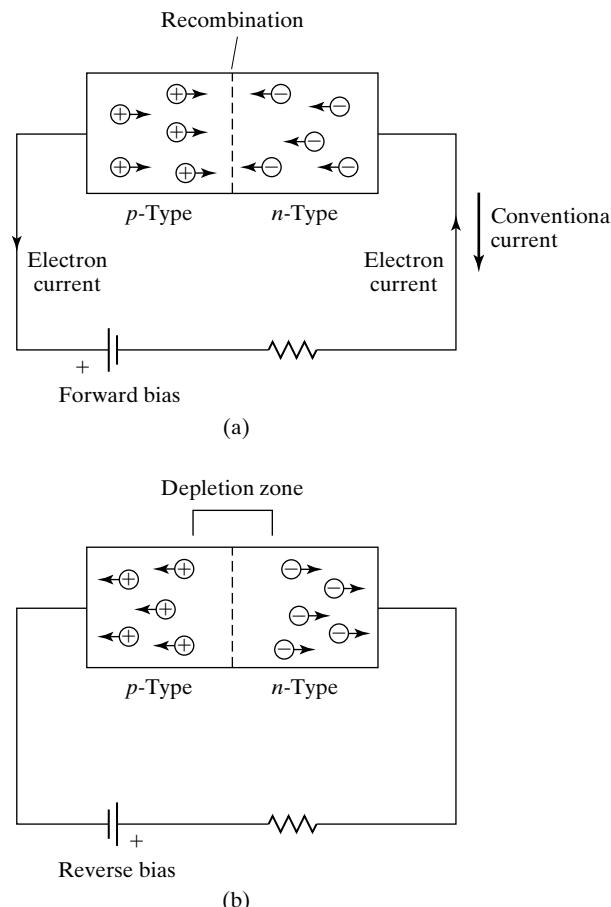
A *pn* junction in equilibrium. In the depletion region the charged bilayer formed by uncompensated acceptor and donor impurities creates a very strong electric field that keeps the region clear of mobile charges.



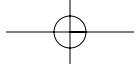
Section 14.3 • The *pn* Junction Diode **461**

Qualitatively, the action of the *pn* diode can be understood as illustrated in Fig. 14.10. In 14.10(a) the diode is connected to an external EMF with the battery's high voltage side connected to the *p* region. In this **forward bias** condition the EMF drives conduction electrons on the *n* side and the positive holes on the *p* side *toward* the *pn* junction. The electrons and holes combine at the interface, producing heat as electrons drop down in energy or, in the case of an LED, light. Thus the diode is continually losing carriers, both *n* and *p*, at the interface. However, the battery drives additional electrons into the diode on the *n* side and hence replenishes the supply of electrons; the battery also pulls electrons out of the *p* side, creating new holes. Thus, the diode's supply of carriers, both *n* and *p*, is continually replenished, and a steady current flows.

In Figure 14.10(b) the external EMF is connected in **reverse bias**, so as to pull the electrons and holes in the diode *away* from the *pn* junction. As the electrons and holes move away from the junction, they increase the width of the depletion zone, in which there are no free charge carriers and there is a strong electric field due to the fixed impurity charges left behind. As the depletion zone grows, this field eventually balances the applied field and the current virtually stops. Thus, the *pn* diode can readily carry current in one direction but not the other.

**FIGURE 14.10**

A *pn* junction diode. In the *n*-type side the mobile charges are electrons; in the *p*-type the mobile carriers are holes. (a) In forward bias the external EMF drives electrons and holes toward the junction, where they recombine. The battery replenishes the supply of carriers (electrons on the *n* side and holes on the *p* side), and a steady current flows. (b) In reverse bias the external EMF tends to draw the carriers away from the junction, widening the depletion zone until the current virtually stops.



462 Chapter 14 • Solids—Applications

We can understand the one-way conduction of the diode in more detail by looking at the band structure diagram again. Figure 14.11(a) shows the band structure in the zero bias condition, that is, with zero externally applied voltage. Because of the built-in electric field of the depletion region, the conduction band in the *n*-type is lower than in the *p*-type by an energy eV_0 . The value of V_0 , called the *contact potential*, depends on the dopant concentration and on the temperature, but it is typically somewhat less than the band-gap energy; for instance, in a silicon diode, $V_0 \approx 0.7$ V. At zero bias, there is no current flow in the diode. However, the situation is not static; at any nonzero temperature, the diode is in a state of dynamic equilibrium in which thermally activated carriers are crossing the junction in both directions. On the *n*-side, some conduction electrons are thermally activated high enough in energy to surmount the eV_0 barrier and *diffuse* to the *p* side.* This **diffusion current** is proportional to the number of electrons activated above the barrier, and so, using (14.1), we can write this current as $I_0 e^{-eV_0/kT}$, where I_0 is a constant. On the *p* side, there is a very small concentration of electrons that are thermally activated from the valence band across the band gap into the conduction band. Any such electrons near the junction will be rapidly swept to the *n* side by the

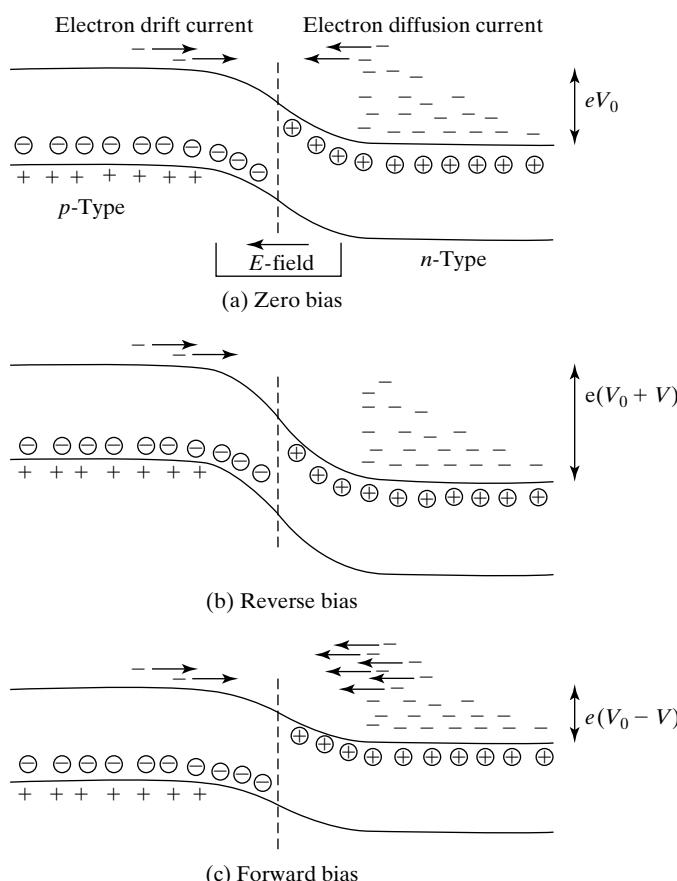


FIGURE 14.11

- (a) In zero bias the net current across a *pn* junction is zero. A thermally activated diffusion current of electrons from *n* to *p* is canceled by a thermally activated drift current of electrons from *p* to *n*. (b) In reverse bias the diffusion current is vanishingly small, and the net current is given by the very small drift current. (c) In forward bias the net current is dominated by a large diffusion current.

*It might seem that these electrons cannot diffuse against the strong electric field in the depletion region; however, recall that these are electrons that have already been activated to high energies — high enough to be above the barrier created by the electric field.

built-in electric field of the depletion region. This charge flow is called a **drift current** (since the E field causes the electrons to drift downstream). Because the net current across the junction is zero, the drift current must be equal in magnitude to the diffusion current, $I_0 e^{-eV_0/kT}$.

In the reverse-bias condition, depicted in Fig. 14.11(b), the externally applied voltage of magnitude V increases the band bending so that conduction electrons on the right now see an energy barrier increased to $e(V_0 + V)$. The diffusion current from right to left is greatly reduced because of the increased activation barrier. However, the drift current due to electrons activated into the conduction band on the left, is unaffected and it now dominates the net current across the junction. We will label the drift current I_R since it dominates in the reverse-biased condition. The drift current is therefore $I_R = I_0 e^{-eV_0/kT}$.

In the forward-bias condition, depicted in Fig. 14.11(c), the externally applied voltage V raises the energy levels on the right so that the energy-level difference is reduced to $e(V_0 - V)$. Because of the lower energy barrier, the diffusion current is increased and is given by $I_0 e^{-e(V_0-V)/kT} = I_R e^{eV/kT}$. This same relation holds for negative applied voltage V in the reverse-biased condition. The net current across the junction is diffusion current minus drift current and is therefore given by

$$I(V) = I_R(e^{eV/kT} - 1) \quad (14.4)$$

This current-voltage relation is graphed in Fig. 14.7. Note that because of the small value of kT at room temperature ($kT_{\text{room}} = 0.025$ eV), any reverse-biased voltage greater than a fraction of a volt will give a diode current dominated by the small drift current $I \approx -I_R$.

We have presented our discussion in terms of the movement of electrons; parallel arguments can be given in terms of the hole movement. For instance, in Fig. 14.11(a) we have shown the small drift current due to conduction electrons on the left; but there is a comparable drift current due to thermally activated valence holes on the right. It is important to note that the suppression of minority carriers plays an essential role here. The small minority carrier population on both sides of the junction results in a very small drift current and a small reverse-biased diode current.

14.4 The Transistor

Perhaps the single most important invention of the twentieth century is the transistor. The transistor is a 3-terminal device that is used as an amplifier or a switch, in which a low-power input signal is used to control a high-power output signal. Transistors come in two basic types: bipolar junction transistors (BJTs) and field-effect transistors (FETs). The BJT was developed first — at Bell Laboratories in 1948 by John Bardeen, Walter Brattain, and William Shockley. For this work, they were awarded the Nobel Prize for physics in 1956. We begin by describing the bipolar junction transistor.

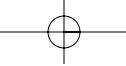
The Bipolar Junction Transistor (BJT)

Bipolar junction transistors can be thought of as two pn junctions back to back; they come in two flavors: npn and pnp . The npn transistor consists of a very thin layer of p -type semiconductor sandwiched between two n -type semiconductors.

John Bardeen
(1908–1991, American)



For the invention of the transistor in 1948, Bardeen (left) and Bell Labs colleagues William Shockley (center) and Walter Brattain (right) were awarded the 1956 Nobel Prize in physics. In 1951 Bardeen joined the faculty at the University of Illinois, where he remained for the rest of his career. In 1955 Bardeen, Leon Cooper, and J. Robert Schrieffer developed the BCS theory of superconductivity, for which they won the Nobel Prize for physics in 1972. Bardeen was the first (and so far, only) person to receive two Nobel Prizes in physics. He was a modest, soft-spoken man. During seminars at Illinois, when he would rise from the audience to ask a question, the room would become very quiet. He was an expert golfer, and on the fairway many of his golf buddies didn't even know that he was a scientist.



464 Chapter 14 • Solids—Applications

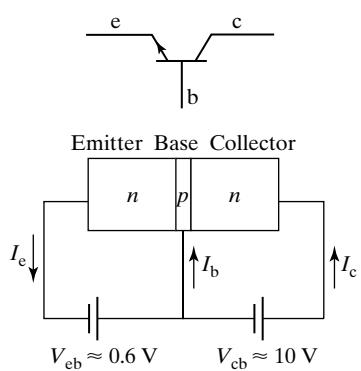


FIGURE 14.12

An *npn* bipolar junction transistor. A very thin *p*-type semiconductor is sandwiched between two *n*-type semiconductors. The emitter-base junction is forward-biased, while the base-collector junction is reverse-biased. The symbol at the top is the standard symbol for a transistor.

As illustrated in Fig. 14.12, the thin central layer, which is usually less than $1\text{ }\mu\text{m}$ thick, is called the **base**, and the two other regions are called **emitter** and **collector**. (The *pnp* transistor has an identical design except *n*- and *p*-regions are reversed. Here we describe the *npn* type only.) When the transistor is used in a circuit, the emitter-base junction is forward-biased by an external voltage $V_{eb} \approx 0.6\text{ V}$ and the collector-base junction is strongly reverse-biased with a large voltage $V_{cb} \approx 10\text{ V}$. The corresponding band structure is shown in Fig. 14.13. Because the emitter-base junction is forward biased, electrons flow from the *n*-type emitter to the *p*-type base. If this was a simple *pn* junction diode, all these electrons would combine with holes in the base. However, because the base region is very thin, most of these electrons diffuse to the collector before they have a chance to recombine with holes. Also, the *p*-type base is very lightly doped to reduce the hole concentration and further decrease the electron-hole recombination rate. Once the electrons reach the collector-base junction, which is strongly reverse-biased, they are swept to the emitter side by the strong electric field in the depletion region. Notice that electrons in the *p*-type base are minority carriers, and so they produce a large drift current from the *p* to the *n* side of the collector-base junction, which would not ordinarily occur in a reverse-biased *pn* junction.

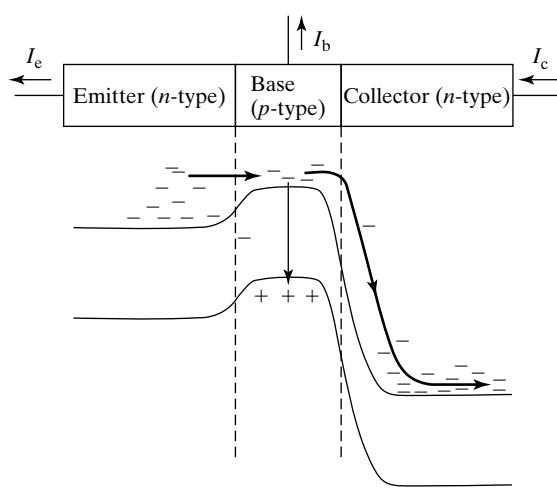
To understand how the *npn* transistor acts as an amplifier, consider what happens when there is a small increase in the base current I_b . Increasing I_b means injecting more holes into the *p*-type base region, making it more positively charged and lowering the electron energy levels. This leads to an increase in V_{eb} , the forward-biased emitter-base voltage. A small increase in the voltage across a forward-biased junction produces a large increase in current since we are on the steeply rising part of the current-voltage diode curve (Fig. 14.7). Thus, a small increase in I_b produces a large increase in the emitter current I_e , which divides between the small base current and the large collector current, $I_c = I_b + I_e$. Over a certain range of currents, the collector current I_c is proportional to the base current I_b ,

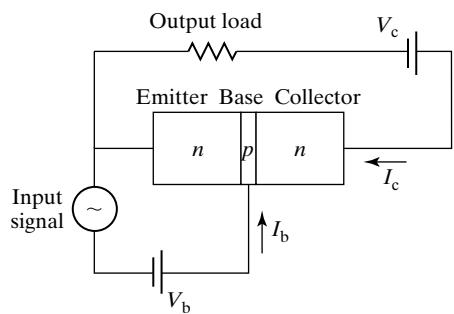
$$I_c = \beta I_b \quad (14.5)$$

where β is a constant that has a typical value of $\beta \approx 100$. A simple amplifier circuit containing a transistor is shown in Figure 14.14.

FIGURE 14.13

Band diagram of an *npn* transistor in operation. The arrows in the lower diagram indicate electron flow. The emitter-base junction is forward-biased, so there is a large electron flow, a diffusion current. There is a small base current due to the small electron-hole recombination rate in the thin base. Most of the conduction electrons in the base diffuse to the collector-base junction, where they are swept to the emitter side by the strong *E* field in the collector-base depletion region. Notice that electron flow is opposite to conventional current flow.



**FIGURE 14.14**

A transistor used as a simple amplifier. A small modulation of the input power (small I_b , small V_b) produces a large modulation of the output power (large V_c , large I_c). The input might be a microphone, and the output a speaker.

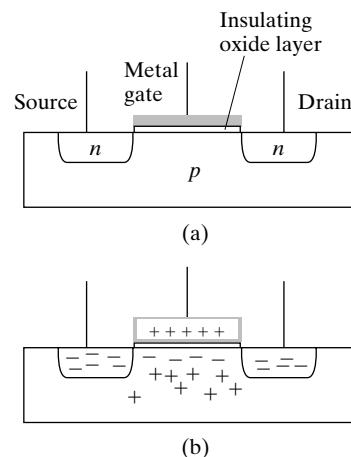
Although it may appear from Figs. 14.12 and 14.13 that the transistor is symmetrical and the emitter and collector are interchangeable, in fact the unsymmetrical operation requires an unsymmetrical construction. Heat is generated at the collector-base junction as electrons slide down the sloping energy ramp, so this side of the transistor requires a good heat sink. Also the emitter and collector are doped differently to optimize their different functions.

The BJT is a superb amplifier; however, its heat dissipation, though small compared to the vacuum tubes it replaced, is still too large to be suitable for use as a switching element in integrated circuits. For this purpose, a very different kind of transistor was developed, the field-effect transistor (FET).

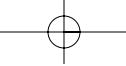
The Field-Effect Transistor (FET)

Field-effect transistors come in many types, and here we describe a particular kind called the metal oxide semiconductor field-effect transistor, or MOSFET. As the name implies, the MOSFET consists of a metal electrode, called the **gate**, separated from a semiconductor (usually silicon) by a thin insulating oxide layer (usually SiO_2). A typical MOSFET has the structure shown in Fig. 14.15(a). Beneath the gate and the oxide is *p*-type silicon, and on either side are *n*-type regions called the **source** and **drain**. This semiconductor structure is somewhat similar to that of an *npn* bipolar junction transistor, with source, gate, and drain corresponding to emitter, base, and collector. The MOSFET differs from the BJT in that the central *p*-type region is not very thin and there is no direct electrical connection to it.

In the MOSFET a voltage applied to the gate has a dramatic effect on the resistance between source and drain. When no voltage is applied to the gate, the electrical resistance between source and drain is quite high because the *npn* structure acts like two *pn* junction diodes back to back. Any voltage difference between source and drain will reverse bias one of the *pn* junctions, and it will not conduct. However, as indicated in Fig. 14.15(b), if a positive voltage is applied to the gate, the positive charge on the gate repels holes and attracts electrons in the *p*-type silicon. The electric field produced by the charge on the gate affects the carrier density in the semiconductor — hence the name, *field-effect* transistor. If the gate voltage is above a threshold value of about +1 V, then in a thin layer below the oxide, conduction electrons outnumber holes and the *p*-type silicon behaves like *n*-type. This so-called **inversion layer** forms a conducting channel between source and drain, lowering the source-drain resistance by several orders of magnitude. The MOSFET thus acts like a switch with the gate voltage controlling the source-drain resistance. The primary advantage of the MOSFET over the BJT is that the input signal (the gate voltage) involves virtually no current, so the input power is extremely low. The low power requirements make the MOSFET an ideal switching element for use in integrated circuits, where self-heating is a significant problem.

**FIGURE 14.15**

(a) A MOSFET (metal oxide semiconductor field-effect transistor). (b) When the gate has a sufficiently large positive charge, holes in the *p*-type semiconductor below the gate are repelled and electrons are attracted, forming a conducting channel (called an inversion layer) between the source and drain. The voltage on the gate controls the resistance between source and drain.



466 Chapter 14 • Solids—Applications

The operation of both BJT and FET transistors is sensitive to temperature. For a transistor to function, the temperature must be high enough that the impurity sites are ionized (so donors produce electrons and acceptors produce holes), but not so high that significant numbers of electrons are activated across the full band gap E_g . Fortunately, room temperature is just about the ideal operating temperature for silicon transistors. But transistors won't function if they overheat, so proper cooling is always important in circuit design. In severe environments, such as in space satellites, the circuit temperature must be carefully regulated.

14.5 Further Semiconductor Applications*

* It seems a pity not to look at some successful applications of semiconductor technology, especially if you have just slogged through the preceding two sections; nevertheless, this section can be skipped if you are pressed for time.

In this section we examine a few other useful applications of semiconductor technology.

LEDs: Light-Emitting Diodes

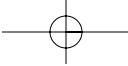
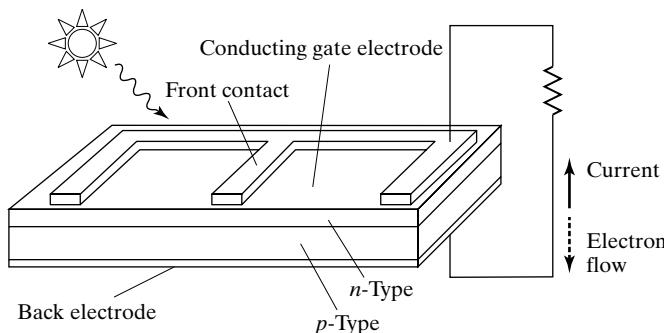
As a current flows in the forward direction through a *pn* diode, electrons are continually meeting holes at the *pn* junction. When an electron drops into a hole, the excess energy can be emitted in the form of a photon. (As mentioned in Section 14.2, this process occurs only in direct-gap semiconductors. See Problem 14.17 for details.) If this photon is in the visible range, the diode will emit visible light. Such diodes, which emit light when a current flows, are called **light-emitting diodes**, or LEDs, and are widely used in the displays of stereos, clocks, and other electrical equipment.

If the current in an LED is large enough, many electrons flow into the *p* region, and before they have a chance to drop into holes, they form an appreciable *population inversion* of excited electrons (Section 11.9). The photon emitted when one electron drops spontaneously into a hole can then stimulate many more electrons to drop and emit further photons. This is the basis of the **diode laser**, whose dimensions can be very small compared to other types of laser and which is finding wide application in fiberoptic communication systems.

Solar Cells

In the LED and diode laser, a current is driven through the device and photons are emitted. It is also possible to reverse this process: If photons shine onto a *pn* diode, they can create electron-hole pairs, which, in turn, cause a current to flow. The energy to drive this current comes, of course, from the incident photons. This is the basis of the **photovoltaic cell**, or **solar cell**, in which light (often sunlight) drives an electric current.

The structure of a solar cell is shown in Fig. 14.16. Thin layers of *p*-type and *n*-type semiconductor (often silicon) are sandwiched between conducting electrodes. The semiconductor layers are quite thin, 10^{-6} m or less, so that light penetrates easily and can be absorbed in the depletion region. When a photon is absorbed within the depletion region, the electron-hole pair created is pulled apart by the built-in electric field, with electrons flowing to the *n* side, and holes flowing to the *p* side. In silicon solar cells, this charge buildup produces a voltage of about 0.5 V, which can drive a current through an external resistor. The voltage produced cannot exceed about 0.5 V because at this level

Section 14.5 • Further Semiconductor Applications **467****FIGURE 14.16**

A photovoltaic cell or solar cell. Electron-hole pairs created near the *pn* junction create a low-voltage current source.

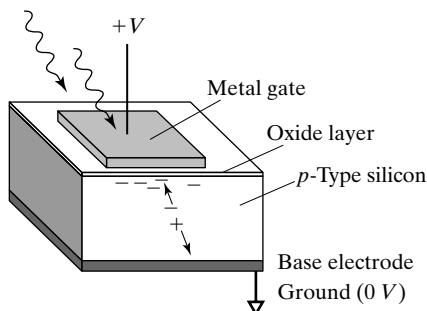
the band bending is reduced to the point that the depletion region collapses and can no longer prevent electron-hole pairs from recombining. However, higher voltages can be attained by placing several solar cells in series.

Commercially available silicon solar cells have an efficiency of about 15%, that is, about 1/6 of the solar power striking the cell is converted to electrical power. The rest of the power is lost to heat. On a clear sunny day, a solar cell with an area of a square meter produces about 100 W of electrical power. (See Problem 14.27.) Efficiencies above 30% have been achieved in some solar cells, and this is close to the theoretical maximum limit set by thermodynamics.

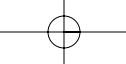
In a device with a structure similar to a solar cell, a *pn* diode can be used as a detector of the particles ejected from radioactive nuclei or other nuclear processes. The diode is maintained with a reverse bias, so that normally no current flows. However, when a nuclear particle passes through the depletion zone, it creates some electrons and holes, which allow a pulse of current to flow through the diode and around the circuit. Thus, each nuclear particle traversing the diode is registered as a pulse of current.

CCD Arrays

A charge-coupled device or CCD is a semiconductor device used to measure the intensity of light. It works by counting photons. As shown in Fig. 14.17, the structure of a CCD is similar to that of a MOSFET, but without a source or drain electrode. A conducting gate electrode is separated from a *p*-type silicon substrate by an insulating oxide. A positive voltage applied to the gate repels holes and attracts electrons. The gate is made of a very thin, semitransparent conductor, usually doped polycrystalline silicon, so that light can penetrate into the absorbing silicon substrate. When a photon with energy greater than the band gap is absorbed in the semiconductor, an electron-hole pair is created. The electric field due to the gate voltage pulls the pair apart, with the electron held near the positive gate and the hole repelled to a ground electrode. Each photon absorbed produces one more electron stored under the gate electrode, so the stored charge is proportional to the number of photons striking the device.

**FIGURE 14.17**

A charge-coupled device or CCD. A photon is absorbed in the *p*-type semiconductor, creating an electron-hole pair. The electron is attracted to the positively charged gate electrode, and the hole is repelled toward the base electrode. The charge stored under the gate is proportional to the number of photons absorbed.



468 Chapter 14 • Solids—Applications



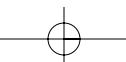
FIGURE 14.18

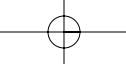
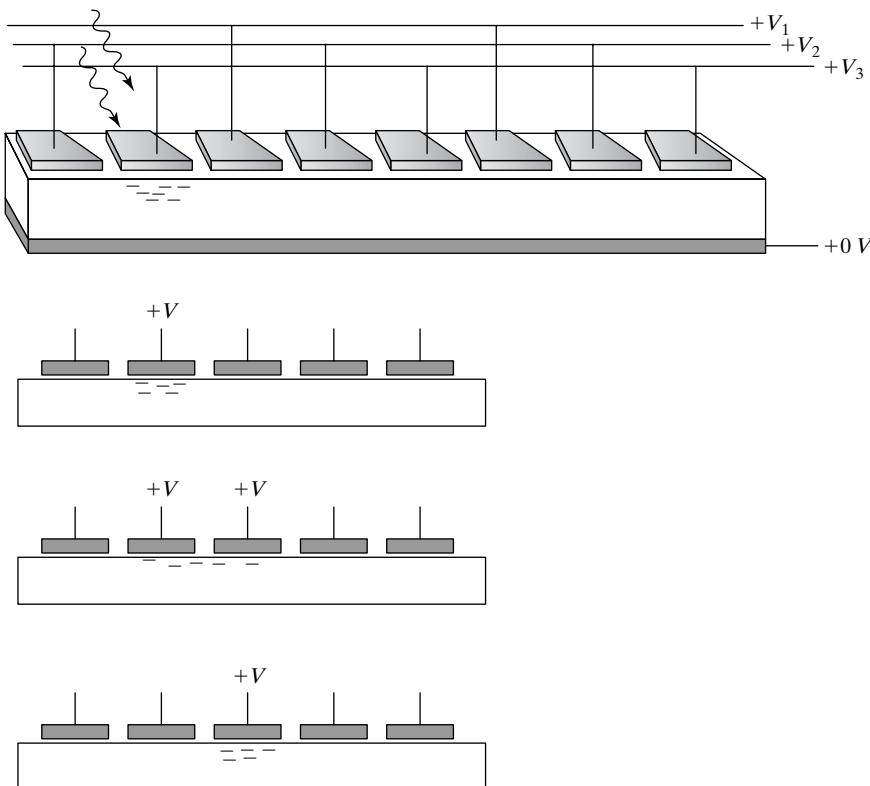
A CCD array chip in a commercial video camera. (a) The central gray rectangle is a 811×508 pixel CCD array with a diagonal length of 8 mm. The individual pixels measure $8 \times 10 \mu\text{m}$ and are too small to be resolved in this picture. (b) A closeup of the upper left corner of the array just barely resolves individual pixels.

Video cameras and digital cameras use **CCD arrays** to record images. A two-dimensional grid or array of CCDs captures the image, with each individual CCD acting as a picture element or **pixel** of the recorded image. In a CCD camera a lens focuses an image onto the CCD array for a controlled exposure time. Exposing the CCD array to light produces a charge in each pixel proportional to the light intensity. Video cameras use array sizes of about 500×500 pixels (standard television resolution), while digital still cameras usually have higher resolution, as high as 2000×3000 pixels. Each pixel in the CCD array is only about 5 to $10 \mu\text{m}$ on a side, so the entire array is usually less than a few cm across, as in Fig. 14.18. Once the image is captured as an array of stored charges, the problem becomes how to read out the charge of each of the 300,000 or more individual pixels. This is accomplished with a kind of “bucket brigade,” as illustrated in Figs. 14.19 and 14.20. The gate voltages are adjusted in a controlled sequence so that the charge packet under each gate is transferred to its neighbor. Each charge packet is passed repeatedly from neighbor to neighbor until it reaches the last gate on the edge of the array, where they are read out serially to an amplifier circuit. Finally, a memory circuit records the charge from each pixel, so that an image can be reconstructed.

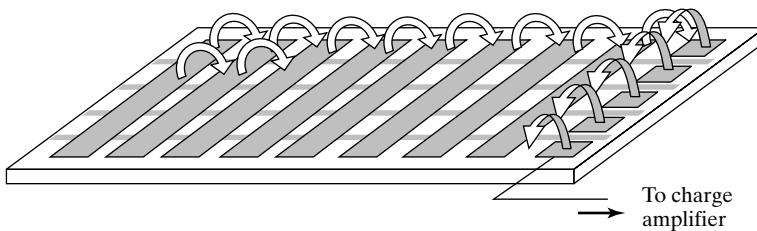
If the CCD camera is operated at room temperature, the image produced is somewhat noisy because thermal excitations as well as photons produce electron-hole pairs. The noise level is usually not severe enough to be noticeable in consumer applications. However, in critical applications, such as astronomical imaging, CCD arrays are cooled to liquid nitrogen temperatures (77 K) in order to reduce thermal noise to negligible levels.

CCD cameras have revolutionized astronomical imaging. Astronomers study images of faint astronomical objects recorded with cameras attached to telescopes. CCDs have vastly improved both the quality and quantity of these



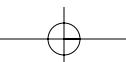
Section 14.5 • Further Semiconductor Applications **469****FIGURE 14.19**

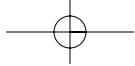
Reading out the charge packets in a CCD array. The charge stored under each gate is transferred to its neighbor by sequentially altering the gate voltages. This bucket brigade passes the charge from gate to gate for eventual readout at the end of the line.

**FIGURE 14.20**

Reading out a two-dimensional CCD array. There is one gate electrode for each column of pixels. Each individual pixel in a column is separated from its neighbors in the same column by controlled doping of the semiconductor (the doping is indicated by the horizontal lines under the gates). Charges are read out serially in a bucket brigade. The charge in each pixel is transferred to its neighbor, first along rows to the right-hand column and then along that column to the electronics that measure the charge.

images. There are two reasons for this: (1) CCD arrays are much more sensitive to light than photographic film, and (2) CCDs provide a quantitative measure of the light intensity in an image. Film records only about 1% of the photons striking it, while CCDs record more than 80% of the incident photons. The result is that astronomers can now capture the light of distant stars nearly 100 times faster than with film. When CCD cameras came into use in the 1980s, the effective light-gathering power of telescopes suddenly increased by a factor of 100. Furthermore, unlike film, CCD cameras have a linear response, meaning that the signal (the magnitude of the charge in each pixel) is exactly proportional to the brightness of the object. This allows a quantitative analysis that was impossible with film.





470 Chapter 14 • Solids—Applications

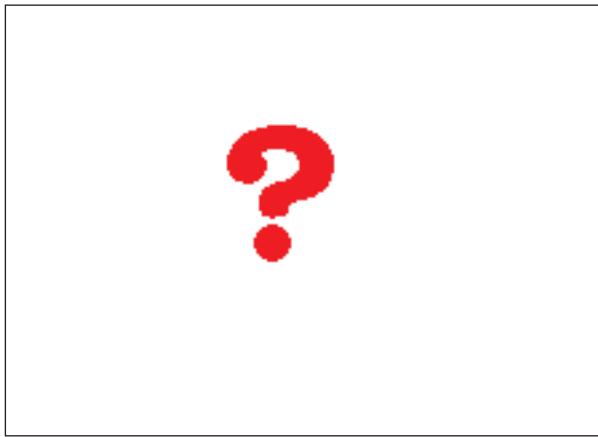
14.6 Integrated Circuits*

* Although of great interest to electrical engineers and others interested in the manufacture of computer chips, the material in this section will not be used again and may be skipped.

Prior to the invention of the transistor, vacuum tubes were used as electronic amplifiers and switches. Compared to transistors, vacuum tubes are large, fragile, and have high power requirements. The first computers, made in the 1940s, were constructed with vacuum tubes and consequently were enormous, cantankerous beasts (see Fig. 14.21). The invention of the transistor was immediately

FIGURE 14.21

The ENIAC (electronic numerical integrator and computer) was one of the world's first electronic computers. Built by the U.S. Army in 1945, it contained 18,000 vacuum tubes, consumed 150 kW of electric power, and had a memory capable of storing twenty 10-digit numbers. Operators programmed the machine with wire plug boards. In 1953 it was upgraded with a magnetic core memory that could store 120 numbers. The ENIAC operated until 1955. It had much less computing power than a modern hand-held programmable calculator.



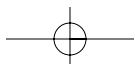
recognized as a technological milestone, and it was thought that the replacement of vacuum tubes with the tiny, robust silicon devices would quickly lead to great advances in technology. But an unexpected problem arose.

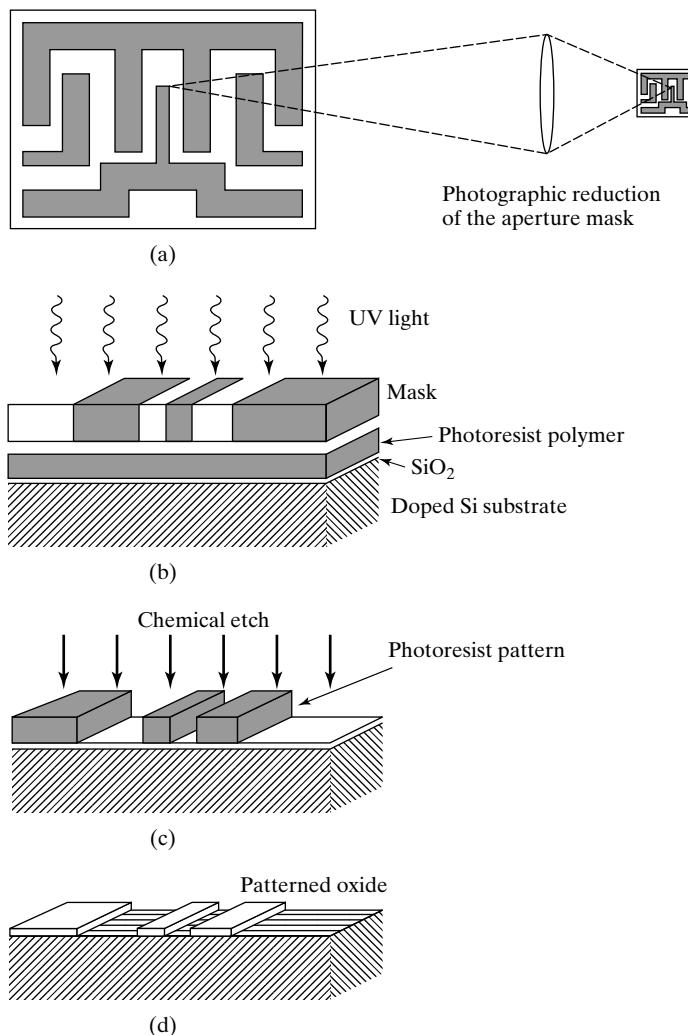
Throughout the 1950s and early 1960s, makers of computers and other electronic circuits faced a growing crisis. Engineers had realized that devices of remarkable power could be built by hooking together great numbers of transistors — even early designs called for many thousands of transistors. However, it is very expensive and time-consuming to wire together thousands of transistors. In addition to the problems of time and cost, there is the even more serious issue of reliability: It is virtually impossible for human workers to make thousands of interconnections without making a mistake somewhere. The problem is really intractable because interconnections cannot be tested individually: Only when all the connections are made can the final device be tested, and then a system failure gives little clue as to which connection is bad. The situation was called “the tyranny of numbers.” Although there existed wonderful designs for computers, it was virtually impossible to construct them.

The solution to the problem was conceived independently by two American engineers: Jack Kilby of Texas Instruments and Robert Noyce of Fairchild Electronics. Both men realized that the key was to manufacture the entire circuit — transistors, resistors, capacitors, and interconnecting wires — all at the same time on a single piece of silicon using **photolithography**. Instead of building the components individually and putting them together later, the components and their connections are built together as a single **integrated circuit**, also called a “chip” (see Figure 14.22).

FIGURE 14.22

An integrated circuit made for use in a computer. This (now obsolete, circa 1988) memory chip can store 4 million bits of information, the equivalent of several hundred printed pages. It can read a single bit in 65 ns or all 4 million in about a quarter of a second. Present-day (2003) memory chips contain about 100 times more memory in the same size package.



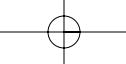
**FIGURE 14.23**

Steps in photolithography.

(a) Mask production. A large pattern defining the circuit layout is photographically reduced to make a mask the size of the final circuit.

(b) UV exposure of photoresist polymer through the mask. Photoresist hardens where exposed to UV light. Sometimes the mask is placed directly against the substrate, as shown here, but often it is held away from the substrate and a demagnified image of the mask is optically projected onto the substrate. **(c)** After the unexposed photoresist is dissolved away, the mask pattern has been transferred to the resist, which protects the underlying oxide from a subsequent chemical etch. **(d)** After the exposed oxide is etched away, the remaining photoresist is removed with a strong solvent. A patterned oxide remains on the Si substrate, ready for the next processing step. Because the feature size is so small (less than 1 μm), the chip manufacturing environment must be exceedingly clean. A single piece of dust anywhere on a lithographic mask will result in dozens of wrecked transistors and a dead chip. The entire assembly line must be in a dust-free clean room, where no humans are allowed, except those wearing isolation suits.

The process of photolithography is illustrated in Fig. 14.23. All circuit structures — transistors, wires, resistors, and so forth — are built in layers on a flat, single crystal of silicon. The process begins by producing an aperture mask containing a layer of the circuit pattern. With microscope optics, a large drawing of the circuit layout is photographically reduced to the final circuit-sized mask. The figure shows how the mask pattern is transferred to a pattern of oxide on the surface of the silicon substrate; similar techniques can produce patterns of metal or doped silicon. A very thin layer of a photosensitive plastic polymer, called **photoresist**, is spread over the substrate. Photoresist (resist, for short) hardens when exposed to ultraviolet light; so after UV exposure through the mask, only the exposed resist remains when the unexposed resist is dissolved away with a chemical solvent. The remaining pattern of hardened resist protects the underlying oxide from a chemical etch, which removes the exposed oxide. After the remaining resist is removed, the pattern of oxide remains, ready for the next processing step. Hundreds of steps involving several



472 Chapter 14 • Solids—Applications



FIGURE 14.24

One of the first integrated circuits built by Robert Noyce at Fairchild Electronics in 1959.



FIGURE 14.25

Edge of an integrated circuit, showing the various layers of circuitry. Each layer is about $1 \mu\text{m}$ thick.

layers are required to produce a complex circuit (see Fig. 14.24 and 14.25). The entire process is automated, and mass-production techniques have kept the cost of computer chips low, even though their complexity and power is increasing exponentially.

Photolithography produces remarkably tiny devices and fantastic numbers of transistors packed into a single chip. The electronics industry is always pushing to produce smaller electronic devices because smaller circuits have shorter signal propagation times, allowing operation at higher frequencies, leading to increased processor speed. Physically smaller computers are faster. Also, smaller transistors require less power, so that the circuit does not overheat as more devices are packed into a smaller area. (See Problem 14.32.) The worldwide demand for more powerful computers is pushing the development of photolithography and integrated circuits faster and further than any industry in human history. In 1965 Gordon Moore, cofounder of Intel, pointed out that the number of transistors on single chips was doubling about every two years. This exponential growth of transistor number, dubbed **Moore's law**, was not expected to continue for very long, but after nearly four decades, Moore's law still continues to hold, as illustrated in Fig. 14.26. Transistor numbers are increasing for two reasons: (1) smaller feature sizes as lithography improves, and (2) larger substrates as materials purity and quality control improves. The Pentium 4 computer chip, produced in the year 2002, has 42 million transistors, 25 layers of circuit patterns, and operates at 2 GHz. This kind of computer power in a package an inch across, costing a few hundred dollars, was simply unimaginable in 1965. In 2001 we didn't have moon bases, but HAL turned out to be small enough and cheap enough that students carry him in their backpacks.*

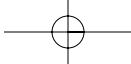
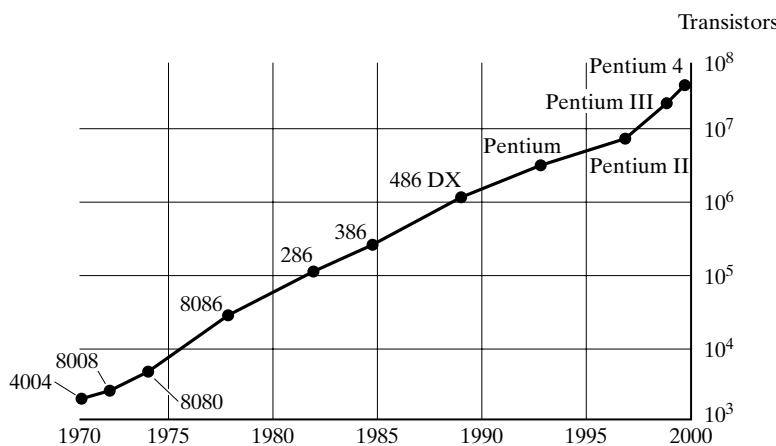
Transistor density on computer chips is limited by the size of the smallest feature that photolithography can produce. The line width of lithographic features is presently limited by diffraction effects and the wavelength of the deep-UV light used (see Problem 14.30). In 2002 the line width in the Pentium chip was $0.18 \mu\text{m} = 180 \text{ nm}$. (A human hair is $50 \mu\text{m}$ in diameter — gigantic on the

Jack Kilby
(born 1923, American)



As a recently hired engineer at Texas Instruments in 1958, Kilby found himself alone in the lab because all his coworkers were on vacation, while he had accumulated no vacation time. During this quiet period, he conceived the integrated circuit. In 1967 he led a team that developed the world's first hand-held calculator. He purposely avoided promotion into an executive position so that he could continue working in the laboratory. He shared the 2000 Nobel Prize in physics for the invention of the integrated circuit.

*We are alluding here to *2001: A Space Odyssey*, the only science fiction film ever made that attempted to be completely faithful to the laws of physics. If you don't know who HAL is, you should see this movie.

Section 14.7 • The Scanning Tunneling Microscope **473****FIGURE 14.26**

Moore's law. A plot of the number of transistors in a computer chip versus time shows a doubling about every 2 years. When will Moore's law fail? At the atomic limit?

scale of transistors in chips.) Sources of extreme UV light and soft X-rays are under development that will produce shorter wavelengths, allowing smaller feature size and the continuation of Moore's law. More exotic technologies, such as using electron beams rather than light to expose the resist, may prolong Moore's law for many more decades, perhaps to the limit of atomic dimensions (Problem 14.33).

14.7 The Scanning Tunneling Microscope*

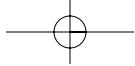
*This section can be skipped without loss of continuity, but you won't regret taking the time to read it.

In 1981 a remarkable new kind of microscope, called the **scanning tunneling microscope** (STM), was invented by Gerd Binnig and Heinrich Rohrer, two researchers at the IBM Zurich Research Laboratory in Switzerland. With the STM, one can routinely image individual atoms on the surface of a metal. True atomic resolution was first achieved much earlier, in 1951, with the invention by Erwin Müller of the **field-ion microscope** — but this is a very specialized device that only allows the imaging of atoms on the tip of an atomically sharp needle made of a refractory (high-melting point) metal such as tungsten or molybdenum. Later, atomic resolution was also achieved with the **transmission electron microscope**, but with this device the sample under examination has to be microscopically thin — a few tens of nanometers or less — so that a beam of high-energy electrons can shine through the sample and project a shadow image onto a fluorescent screen. With the STM, one could for the first time, routinely image atoms on the surface of any electrically conducting sample — either metal or semiconductor. The STM also turned out to be remarkably inexpensive and easy to make, and so it quickly became a standard analytical tool used in thousands of laboratories around the world. The invention of the STM spawned a whole new family of instruments called scanning probe microscopes. One of these, the atomic force microscope (AFM), allows imaging of electrically insulating samples with near-atomic resolution. For the invention of the STM, Binnig and Rohrer shared the Nobel Prize in physics in 1986 with the German physicist Ernst Ruska, who pioneered development of the electron microscope.

Robert Noyce
(1927–1990, American)



A man of wide interests, Noyce was a natural leader as well as a brilliant engineer. While working at Fairchild Semiconductor, Noyce independently conceived the idea of integrated circuits, just a few months after Kilby. In 1968, Noyce and colleague Gordon Moore (of Moore's law) founded Intel Corporation, soon to become the world's largest maker of computer chips. Noyce was much admired by colleagues and employees as an inspiring manager. At his insistence, executives ate lunch and shared office space with the engineers, lest they cease to breathe the air of innovation. In 1990, a vigorous Noyce died prematurely of a sudden heart attack. The Nobel prize is not awarded posthumously.



474 Chapter 14 • Solids—Applications

To understand the operation of the STM, we must review the concept of the **work function** ϕ of a metal, which was covered briefly in Section 4.3. The work function is the binding energy of a conduction electron in a metal. Although the conduction electrons can move about freely *within* the metal, actually *removing* an electron from the metal requires about 4 or 5 eV of energy — the exact value depends on the metal. The values of the work function in some metals are listed in Table 14.2; we will take as a typical value, $\phi = 4$ eV.

TABLE 14.2

Work functions of selected metals. Most metals have a work function of about 4 or 5 eV. Alkali metals such as cesium have unusually low work functions.

Metal	ϕ (eV)
Ni	5.2
Au	5.1
Cu	4.7
W	4.6
Fe	4.5
Sn	4.4
Al	4.3
Ag	4.3
Cs	2.1

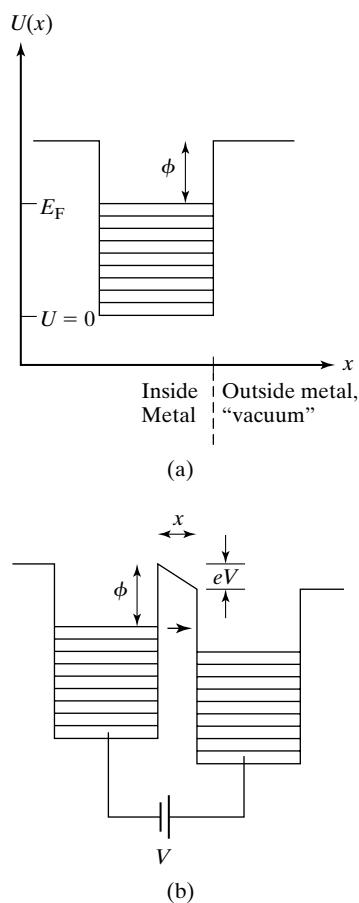


FIGURE 14.27

(a) Energy-level diagram for electrons in a metal. The electrons fill the quantum states up to the Fermi level E_F , which is a work function ϕ below the “vacuum level,” the energy of the electrons at rest outside the metal. **(b)** Electrons can tunnel between two metals separated by a distance x , if x is small enough and there is a voltage difference V between the metals.

An energy-level diagram of the electrons in a metal is shown in Fig. 14.27(a). Conduction electrons in a metal behave almost as free electrons, so the potential $U(x)$ that the electrons see may be approximated as a simple square well of finite depth. The quantum states of this well are filled with electrons (two per level, in accordance with the Pauli principle) up to a maximum filled energy level, the Fermi level E_F . The Fermi level lies one work function below the “vacuum level,” the energy of a stationary ($KE = 0$) free electron far outside the metal.

When two metal surfaces are brought near each other, some distance x apart, and a voltage difference V is established between the metals by attaching them to a battery, the energy-level diagram is as shown in Fig. 14.27(b). The tilted vacuum level between the two metals represents the varying potential energy of an electron in the electric field E between the two metals: $U(x) = -eV(x) = -eEx$. Because of the voltage difference V between the two metals, the Fermi level of the left metal is raised an energy eV with respect to the Fermi level on the right (here, eV means “charge e times voltage V ,” not “one electron-volt”). Electrons near the Fermi level in the left metal are now at the same energy as unoccupied states in the right metal, and they can tunnel through the potential barrier between the metals if the barrier thickness x is sufficiently small.

In a scanning tunneling microscope a sharp metal needle, called the “tip,” is brought very close to the conducting surface to be imaged, as in Fig. 14.28, and a voltage difference V is applied between the tip and the sample. When the tip is within about one nanometer of the surface, electrons begin to tunnel between the tip and the sample, producing a tunneling current I . This tunneling current depends exponentially on the distance z between the tip and

Section 14.7 • The Scanning Tunneling Microscope 475

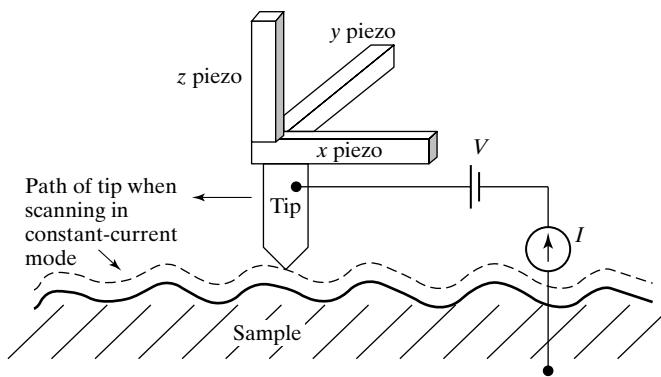


FIGURE 14.28
Operation of a scanning tunneling microscope (STM).

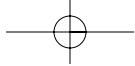
the sample and varies as

$$I(z) \propto \exp(-2z/\lambda) \quad (14.6)$$

where λ is a length called the tunneling depth, which is a measure of the distance that the wave function leaks into the classically forbidden region. In the next subsection we will derive (14.6) and show that the tunneling depth has the value $\lambda \approx 0.1$ nm. According to (14.6), the tunneling current provides an exquisitely sensitive measure of the height z of the tip above the sample. When z is increased by a distance $\lambda \approx 0.1$ nm, the current $I(z)$ decreases by a factor of $e \approx 2.7$; when z is increased by a distance of 2λ , the current decreases by a factor of $e^2 \approx 7.4$, etc.

So how do we use the tip and the tunneling current to form an image of the sample surface? The tip is held in place above the sample by piezoelectric ceramic rods. A piezoelectric ceramic ("piezo," for short) is a material whose dimensions change slightly when an electric field is applied. For instance, one can buy a 2-cm-long piezo rod that lengthens by 10 nm when a voltage difference of $\Delta V = 1$ V is applied between its ends. If you apply 2 V, the rod stretches by 20 nm; if you apply $\Delta V = -1$ V, the rod shrinks by 10 nm. By attaching the tip to a triad of such piezo rods, as in Fig. 14.28, and applying appropriate voltages to the rods, one can move the tip around in the x , y , and z directions with subnanometer control. Time-varying voltages applied to the x and y piezos cause the tip to *raster* over the surface, that is, scan a square area, like the electron beam in a TV tube. As the tip scans the surface, the height is continuously adjusted to keep the tunneling current constant, causing the tip to ride up and down over the hills and valleys of the surface. (In this case the "surface" is a contour of constant electron wave-function amplitude.) A computer records the height of the tip, which is proportional to the z voltage, as a function of its x - y position and uses this information to construct an image of the surface.

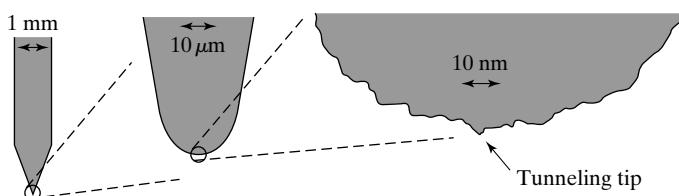
This scanning technique will provide atomic resolution only if the tip is atomically sharp — sharp enough for the tip to fit down inside the very slight wave-function depressions between atoms. To the great surprise of the inventors, producing atomically sharp tips turned out to be easy. If a needle is sharpened by some mechanical or chemical process and then viewed with a microscope, the tip almost always *appears* to be rather blunt, with a radius of curvature of many hundred nanometers (see Fig. 14.29). However, when viewed on an atomic scale, the metal surface is almost always rough and there



476 Chapter 14 • Solids—Applications

FIGURE 14.29

A tunneling tip viewed at three different magnifications. An uncontrolled microfeature often dominates the tunneling in an STM.



are always *uncontrolled microfeatures*, one of which is likely to stick out from the surface just a fraction of a nanometer farther than the others. Because of the smallness of the tunneling depth λ , this microfeature will dominate the tunneling current and produce images with atomic resolution, such as in Figs. 14.30 and 14.31.

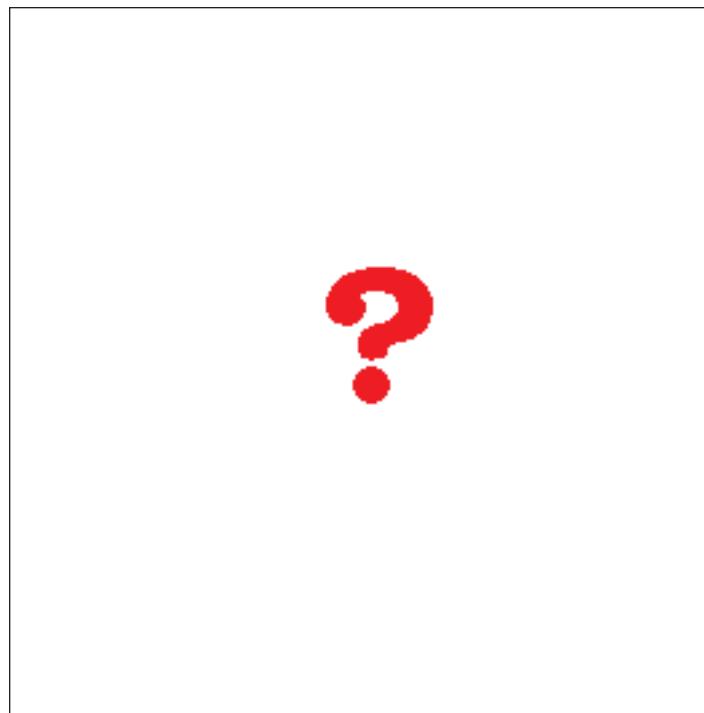
FIGURE 14.30

Atomic resolution STM image of the surface of a silicon crystal. The two nearly vertical lines are atomic steps separating terraces. The field of view is $30 \text{ nm} \times 60 \text{ nm}$.



FIGURE 14.31

A quantum corral under construction. These four STM images show individual iron atoms on the flat surface of a single crystal of copper. The STM tip not only images the surface, but is also used to nudge the atoms about, slowly constructing a circle of 48 iron atoms. The electronic wave function in the interior of the corral shows distinct circular ripples — a consequence of confining electrons in a circular potential.



The Tunneling Depth λ

In this subsection we derive an expression for the tunneling depth λ , described above. This will be a review of material in Section 7.10, but the notion of tunneling is so central to an understanding of quantum mechanics that it is worth reviewing. As seen in Fig. 14.32, we consider an electron of energy E in a well of finite depth U_0 , where the zero of energy is taken to be at the bottom of the well and $E < U_0$. This is a good model for a conduction electron at the Fermi level in a metal with a work function $\phi = U_0 - E$. We set the origin of the x axis at the right edge of the box. As explained in Section 7.8, the electron wave function $\psi(x)$ is a sinusoidal function inside the box, while outside the box in the classically forbidden region, the wave function decays exponentially. We now re-derive this exponential decay.

Beginning with Schrödinger's equation,

$$\psi''(x) = \frac{2m}{\hbar^2}(U(x) - E)\psi(x) \quad (14.7)$$

we consider the region $x > 0$, outside the box, on the right side. In this region we have $U(x) - E = U_0 - E = \phi$, and (14.7) becomes

$$\psi''(x) = \left(\frac{2m\phi}{\hbar^2}\right)\psi(x) = +\alpha^2\psi(x) \quad x > 0 \quad (14.8)$$

where $\alpha = \sqrt{2m\phi/\hbar^2}$. The physically allowed solution to this equation is the exponentially decaying function

$$\psi(x) = \psi_0 e^{-\alpha x} \quad (14.9)$$

where $\psi_0 = \psi(x = 0)$. (There is another mathematical solution to (14.8) given by $\psi(x) = \psi_0 e^{+\alpha x}$, but we throw out this solution on physical grounds: This solution *diverges* exponentially, so the wave function cannot be normalized.) We now define the tunneling depth λ as

$$\lambda = \frac{1}{\alpha} = \frac{\hbar}{\sqrt{2m\phi}} \quad (14.10)$$

and we can rewrite equation (14.9) as

$$\psi(x) = \psi_0 e^{-x/\lambda} \quad (14.11)$$

The tunneling depth λ is a measure of how far the wave function “leaks” into the classically forbidden region.

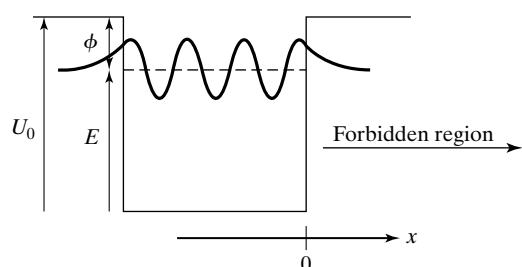
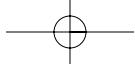


FIGURE 14.32

Bound state wave function in a finite potential well (nonrigid box).



478 Chapter 14 • Solids—Applications

Example 14.2

What is the value of the tunneling depth in a typical metal with a work function of $\phi \approx 4$ eV?

From equation (14.10) we have for the tunneling depth:

$$\lambda = \frac{\hbar}{\sqrt{2m\phi}} = \frac{1.05 \times 10^{-34}}{\sqrt{2 \times (9.1 \times 10^{-31} \text{ kg}) \times (4 \text{ eV}) \times (1.6 \times 10^{-19} \text{ J/eV})}}$$

$$\approx 1 \times 10^{-10} \text{ m} = 0.1 \text{ nm} \quad (14.12)$$

This calculation confirms the value of λ , which we used earlier when discussing tunneling tips. Since all metals and semiconductors have nearly the same size work function (within a factor of 2 or so), the tunneling depth is always about 0.1 nm. So atomic resolution can be achieved on any conducting surface.

We are now in a position to prove the relation (14.6). A tunneling current — a flow of tunneling electrons — can occur when a potential barrier separates two wells, as in the case of two metals near each other [Fig. 14.27(b)]. Consider an electron in a single potential well of finite depth. A solution to Schrödinger's equation is shown in Fig. 14.32. Now imagine bringing a second empty well very near (but not touching) the first, much as in Fig. 14.27(b). The electron is no longer in a single potential well, but is now in the left half of a double well. In the vicinity of the barrier between the wells, the wave function for the electron now appears as shown in Fig. 14.33. The probability that the electron, initially on the left of the barrier, will tunnel through to the right side is proportional to the square of the wave function (14.11) on the right,

$$|\psi(x = d)|^2 = |\psi_0 e^{-d/\lambda}|^2 \propto e^{-2d/\lambda}$$

Thus, the tunneling current is proportional to this tunneling probability, and we have established (14.6).

As mentioned before, one drawback of the STM is that it can be used only to image clean, unoxidized conducting surfaces. Insulators or metals with thick oxide layers cannot be imaged with an STM because a tunneling current can occur only between electrical conductors. Another device, the atomic force microscope (AFM), developed a few years after the STM, does allow imaging of insulating surfaces; however, it does not routinely achieve atomic resolution.

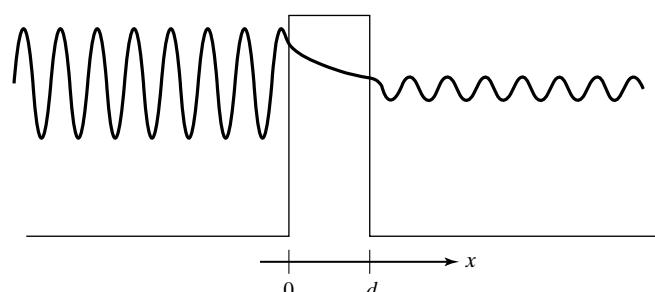
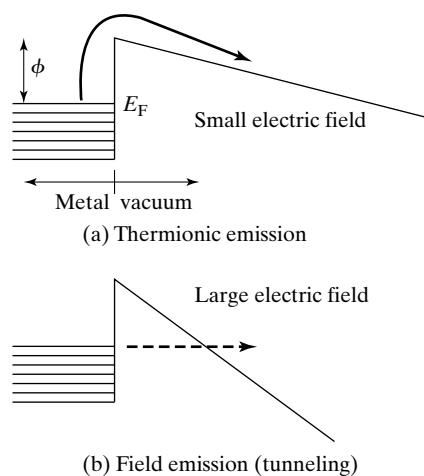


FIGURE 14.33

An electron initially in the left half of the potential barrier can tunnel into the right half if the barrier is thin enough and not too high.

**FIGURE 14.34**

Thermionic emission versus field emission from a metal surface.

- (a) In thermionic emission, electrons at the Fermi level in the metal are thermally activated over the work function energy barrier. (b) In field emission the external electric field is so strong that the barrier becomes thin enough for electrons to tunnel through.

More About the Work Function

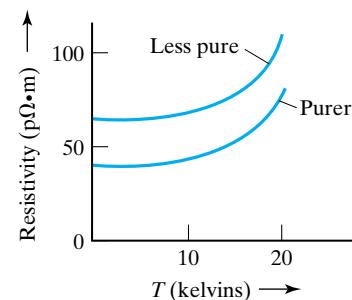
The notion of the work function is central to an understanding of two important phenomena in metals: thermionic emission and field emission. When a metal is held at a negative voltage with respect to ground and then heated, electrons literally boil off the surface, and this process is called **thermionic emission**. The situation is depicted in Fig. 14.34(a). Because of the applied negative voltage, the electron energy levels in the metal are raised and the electron energy outside the metal slopes away, reflecting the existence of an electric field surrounding the metal. At high temperatures electrons at the Fermi level are thermally activated up over the work function energy barrier, and once over the barrier, they are carried away by the electric field, producing a current given by $I \propto e^{-\phi/kT}$. The electron beam in a television tube is generated by thermionic emission from a heated tungsten filament.

As shown in Fig. 14.34(b), if a very large negative voltage is applied to the metal and the resulting electric field outside the metal is very strong, the energy barrier can become thin enough for electrons to tunnel through to the vacuum. This process is called **field emission**, and it occurs even if the metal is cool. The field-emission current varies exponentially with the electric field E as $I \propto \exp(-\kappa\phi^{3/2}/E)$, where κ is a constant (see Problem 14.40). Field emission is used in some electron microscopes and in other applications where electron beams with very high current are needed.

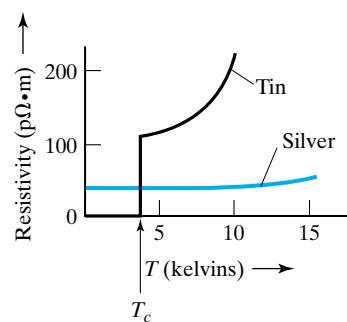
14.8 Superconductivity*

*This section can be skipped without loss of continuity, but it's good stuff.

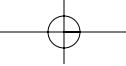
In Section 13.7 we saw that at high temperatures the electrical resistivity of metals is due mainly to electron–phonon collisions, that is, collisions between the conduction electrons and vibrations of the lattice. As the temperature is lowered, electron–phonon scattering decreases and eventually the resistivity is dominated by electron–impurity scattering, and so purer samples exhibit a lower resistance at low temperatures, as shown in Fig. 14.35. However, as seen in Fig. 14.36, when the temperature falls below a critical value, some metals undergo a sudden transition to a state of zero resistance (or infinite conductivity) called **superconductivity**. The temperature at which this transition occurs

**FIGURE 14.35**

The resistivities of two samples of silver, as functions of temperature. The purer sample has lower resistivity.

**FIGURE 14.36**

The temperature dependence of the resistivity of tin, a superconductor. At the critical temperature $T_c = 3.7$ K, the resistivity drops abruptly to zero. Below T_c the resistivity is exactly zero, and we say that the material is superconducting. Above T_c the sample is said to be in the normal state. For comparison, the colored curve shows the behavior of silver, which remains in the normal state at all temperatures.



480 Chapter 14 • Solids—Applications

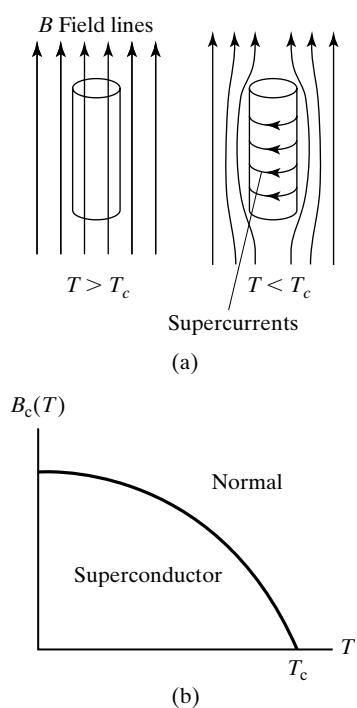


FIGURE 14.37

- (a) Behavior of a type I superconductor in a magnetic field.
- (b) Critical field B_c versus critical temperature for a type I superconductor.

is called the **critical temperature** T_c , and it is usually only a few degrees above absolute zero. Ordinary non-superconducting metals are called **normal** metals, and superconducting metals above the critical temperature are said to be in the **normal state**. Many alloys and about a third of the metals in the periodic table are superconductors.

Superconductivity was discovered in 1911 by the great Dutch experimentalist H. Kamerlingh Onnes. In 1908 Onnes had developed refrigeration techniques allowing the production of liquid helium, which boils at 4.2 K. His was then the only laboratory in the world that could produce such low temperatures, and he began a systematic study of the low-temperature properties of metals. In 1911 he measured the temperature dependence of the resistivity of mercury (Hg) (chosen because it could be easily purified). Onnes was astounded to discover that when the temperature was lowered to 4.15 K, the sample resistance dropped suddenly to unmeasurably low values.

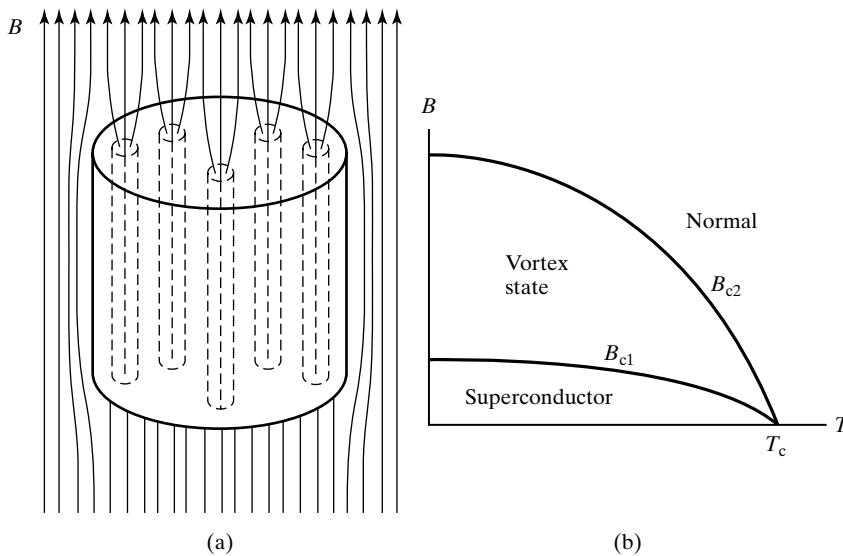
In addition to zero resistivity, superconductors have another remarkable defining characteristic: They expel magnetic fields. When a superconducting material is placed in a magnetic field and is cooled below its critical temperature, the magnetic field is expelled from its interior. This occurs because superconducting currents (supercurrents) appear on the surface of the superconductor, creating a magnetic field that exactly cancels the externally applied field inside the sample. This is called the **Meissner effect** and is illustrated in Fig. 14.37(a). Magnetic fields tend to suppress superconductivity, and as shown in Fig. 14.37(b), there is a temperature-dependent critical magnetic field $B_c(T)$, above which superconductivity is destroyed completely.

Superconductors are classified as either **type I** or **type II**, depending on their behavior in a magnetic field. Type I superconductors behave as shown in Fig. 14.37. Below the critical field B_c , the expulsion of the magnetic field is complete, and above it, the superconductivity is completely destroyed. All of the superconducting elements and some alloys are type I. As shown in Table 14.3, the type I superconductors all have low critical temperatures ($T_c < 10$ K) and low critical fields ($B_c < 0.2$ T), making them unsuitable for high-magnetic field applications.

TABLE 14.3

Critical temperatures and critical fields of some conventional “*lo-T_c*” superconductors. Critical fields measured at $T = 0$ in units of teslas.

	Material	$T_c(K)$	$B_c(0)$ or $B_{c2}(0)$ (in teslas)
Type I	W	0.012	0.00011
	Zn	0.88	0.0053
	Al	1.17	0.0105
	In	3.41	0.0293
	Hg	4.15	0.0411
	Pb	7.19	0.0803
	Nb	9.25	0.198
Type II	PbMoS	14.4	6.0
	NbN	16.0	15.3
	Nb ₃ Sn	18.1	24.5
	Nb ₃ Ge	23.2	34.0

**FIGURE 14.38**

Magnetic behavior of a type II superconductor. (a) In the vortex state the magnetic field penetrates the superconductor in thin filaments. Each filament is surrounded by a solenoid of shielding supercurrent, which encloses a thin tube of normal material called a vortex tube. (b) The B versus T plane.

Type II superconductors have a more complex behavior, illustrated in Fig. 14.38. In these superconductors there are two critical fields, labeled B_{c1} and B_{c2} . Below B_{c1} the field is expelled completely, as in type I superconductors. However, above B_{c1} the magnetic field begins to penetrate the sample without destroying the superconductivity. The field penetrates only in narrow filaments, called **vortex cores**. Within the vortex cores the sample is normal, but between the cores the sample remains superconducting. As the field is increased, the vortex cores become more densely packed and a greater fraction of the sample is normal until, at the upper critical field B_{c2} , the vortex cores overlap and the sample becomes completely normal. Many superconducting alloys are type II; a few are listed in Table 14.3. Note that type II superconductors have higher critical temperatures than type I. Type II superconductors are of great technological interest because the upper critical field B_{c2} is usually quite high, typically tens of teslas, making type II superconductors ideal candidates for high-field magnet wire.

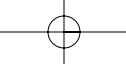
It should be noted that because of the Meissner effect, a superconductor is more than simply a perfect conductor. A perfect conductor has zero resistance, and consequently, the electric field E in perfect conductors is always zero. (Otherwise, there would be an infinite current.) From Faraday's law we know that any change in the magnetic field B necessarily produces an electric field E . Therefore, if E is always zero, it follows that B cannot change; that is, B must be constant in a perfect conductor. However, in a superconductor, B is not merely constant, it is zero.

The distinction between perfect conductors and superconductors can be seen using two versions of the magnetic levitation experiment shown in Fig. 14.39. Suppose that we place a superconductor in a field-free region

FIGURE 14.39

A magnet is levitated above a sample of the hi- T_c superconductor yttrium–barium–copper oxide ($\text{Y}_1\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$) cooled in a dish of liquid nitrogen. The magnet can either be lowered toward a cold superconductor, already below T_c ; or it can be placed on a warm superconductor, in the normal state, which is then cooled below T_c . The magnet is levitated in both cases. In the first case the experiment demonstrates Faraday's law; in the second, the Meissner effect.





482 Chapter 14 • Solids—Applications

(so that $B = 0$) and then bring up a magnet. By Faraday's law, the approaching magnet induces currents in the superconductor, and the magnetic field of these currents holds up the magnet. Since the material is superconducting, the currents persist, and the magnet remains levitated indefinitely. Now suppose that the magnet is placed *on* the superconductor at a temperature above T_c and then the system is cooled. As the temperature passes T_c , currents appear in the superconductor (to eject the B field) and these currents lift the magnet up and hold it levitated above the superconductor. This is the Meissner effect.

The BCS Theory of Superconductivity

Superconductivity was discovered in 1911, but it was not until 1957 that a satisfactory theory of superconductivity was found. The long delay between discovery and explanation is an indication of how difficult it was to formulate the theory. Some of the greatest physicists of the twentieth century, including Landau and Feynman, attempted to solve the problem, without success. But in 1957 three American physicists at the University of Illinois, John Bardeen, Leon Cooper, and J. Robert Schrieffer, cracked the problem and published what is now called the BCS theory of superconductivity. Details of BCS theory involve quantum calculations well beyond the level of this text, so we can present only a qualitative description.

According to BCS theory, superconductivity arises from a subtle interaction between the conduction electrons and the vibrations of the lattice (phonons). Prior to BCS, there were two clues that superconductivity somehow involved phonons:

1. It was observed that very good conductors, such as copper, silver, and gold, are not superconductors. Only rather poor conductors, like lead and niobium, which have a high room-temperature resistivity because of strong electron–phonon scattering, exhibit superconductivity.
2. It was found that varying the relative abundance of different isotopes of a superconductor has an effect on the critical temperature. Samples with lighter nuclei (fewer neutrons) have a higher T_c . For instance, in mercury, when the mean atomic mass is varied from 203.4 u to 199.5 u, the critical temperature rises from 4.146 K to 4.185 K. A smaller nuclear mass means larger amplitude lattice vibrations at a given temperature, and so this **isotope effect** is further evidence that superconductivity is somehow tied up with the electron–phonon interaction.

BCS theory predicts that the conduction electrons in a superconductor form bound pairs, called **Cooper pairs**. The mechanism of Cooper-pair formation is depicted in Fig. 14.40. When a negatively charged conduction electron passes through the lattice of positively charged ion cores, it pulls the ions toward it, leaving a wake of net positive charge. Two electrons passing in opposite directions are each attracted to the positive wake of the other. If the temperature is low enough so that random thermal motion does not disrupt

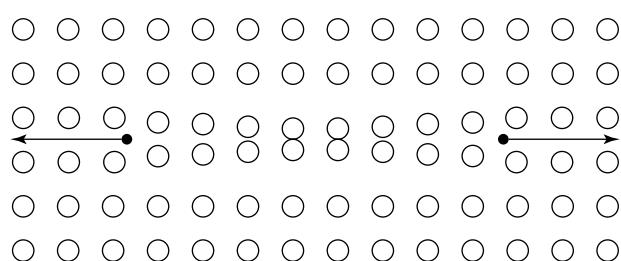


FIGURE 14.40

Cooper-pair formation. When a conduction electron passes through the lattice, it leaves a net positive charge in its wake. Two conduction electrons traveling in opposite directions are each attracted to the positively charged wake of the other.

the wake, and the electron–phonon interaction is large enough to produce a strong wake, then the attraction to the wake can actually be greater than the electron–electron Coulomb repulsion, and the two electrons form a bound pair. BSC theory predicts that the binding energy of the Cooper pair, called the **gap energy** E_g , has a value at zero temperature given by

$$E_g = 3.5kT_c \quad (14.13)$$

It turns out that once the Cooper pairs have formed, lattice imperfections and vibrations can scatter them only by dissociating them, which requires an energy given by (14.13), which is of order 10^{-3} eV. Below the critical temperature, thermal energies are insufficient to do this. Therefore, the Cooper pairs cannot be scattered by the lattice, and the resistivity of the solid is zero.

BCS theory describes superconductivity as a kind of Bose–Einstein transition (Section 13.12). The two electrons of a Cooper pair have equal and opposite momenta and they also have opposite spin; one electron is spin up and the other spin down. Consequently, the bound Cooper pair is a spin-zero object and is a boson. The size of the Cooper-pair wave function is quite large, typically a few hundred nanometers in diameter (see Problem 14.45), so Cooper pairs overlap each other. As described in Section 13.12, when identical bosons overlap, a large fraction of the bosons condense into the same quantum state — a Bose–Einstein transition. Because all of the Cooper pairs are in the same quantum state, their motions are very strongly correlated. It is this lockstep motion of the electrons in the Bose–Einstein condensate that prevents scattering by the lattice and produces zero resistivity.

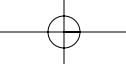
High-Temperature Superconductivity

The potential applications of materials with zero resistance are obvious and numerous. (For example, there would be no joule heating losses in power lines of zero resistance.) We will describe a few successful applications shortly, but the need to keep superconducting circuits at liquid helium temperatures has meant that all such applications are expensive and specialized. Liquid helium costs about \$5 per liter (the cost of not-very-good wine) and requires specialized equipment to handle. For this reason, there was great excitement in 1986 when the two physicists J. Georg Bednorz and K. Alex Müller at the IBM laboratory in Zurich, Switzerland, discovered the first of a class of superconductors with higher critical temperatures. These new **high-temperature (hi- T_c) superconductors** are layered ceramic materials with planes of copper oxide. Some have critical temperatures higher than 100 K. (See Table 14.4.) Since

TABLE 14.4

Critical temperatures of some hi- T_c superconductors. The HgTlCaBaCuO compound, discovered in 1995, has the highest known critical temperature at atmospheric pressure, as of the year 2002. Higher T_c s have been reported, but only in samples held under extremely high pressure.

Material	T_c (K)
$\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$	30
La_2CuO_4	40
$\text{YBa}_2\text{Cu}_3\text{O}_7$	92
$\text{Tl}_2\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$	125
$\text{Hg}_{0.2}\text{Tl}_{0.8}\text{Ca}_2\text{Ba}_2\text{Cu}_3\text{O}$	139



484 Chapter 14 • Solids—Applications

liquid nitrogen has a temperature of 77 K and is inexpensive (about \$0.30 per liter), the new materials can be maintained in their superconducting state relatively easily and cheaply. Therefore, there has been a tremendous effort to find ways to manufacture the new superconductors conveniently and to discover still more of them with even higher critical temperatures. The hi- T_c materials are all type II superconductors and have extremely high upper critical fields; often, B_{c2} is hundreds of teslas, higher than can be measured directly. Producing magnet wire with the new materials has been quite difficult because the hi- T_c materials are all very brittle — their mechanical properties are quite similar to that of a coffee mug — but substantial progress has been made, and practical hi- T_c magnet wire is now being made.

The superconductivity of the new high-temperature superconductors is believed to be due to the formation of electron pairs, as in the BCS theory. However, in this case the pairs are apparently *not* caused by the same electron-lattice interaction as in the BCS theory. Just what does cause the formation of pairs is not known and remains the subject of intense research.

Applications

The complete disappearance of resistance and the resulting absence of joule heating suggest several obvious uses for superconductors. Perhaps the most successful application to date has been to use superconducting wire to carry the large currents needed in high-field electromagnets. Such superconducting magnets are widely used in modern particle accelerators (Section 18.11). They are also used to produce the stable, large-volume magnetic fields needed for magnetic resonance imaging in medicine (Section 9.8).

Another application under study is the reduction of losses in power lines that deliver electricity from power plants. At present, the cost of refrigerating superconducting wires outweighs the potential saving, but the new hi- T_c superconductors may make such applications feasible in the future.

Another area of study is at a much smaller scale. The miniaturization of computers is limited by the heat generated in the conductors linking the semiconductor elements. (A high density of these elements implies a high density of heat generation, resulting in excessive temperatures.) Use of superconducting devices and connections could eliminate this problem.

All the applications mentioned so far take advantage of the superconductor's zero resistance. A possible application of the magnetic properties of superconductors is to low friction bearings whose surfaces could be held apart by the Meissner effect, as in the levitation shown in Fig. 14.39.*

Still another class of applications of superconductivity involves the *Josephson effect*. This quantum phenomenon causes an alternating current to flow between two superconductors separated by a thin insulator when a small DC voltage is applied. Since the frequency of the alternating current is very sensitive to the applied voltage and to any applied magnetic field, the Josephson effect can be used to measure voltages accurately and to measure extremely small magnetic fields (the latter using the so-called *superconducting quantum interference device*, or SQUID). The Josephson effect is also the basis for several high-precision measurements of the fundamental constants.

*An application that is frequently mentioned is the levitation of trains by superconducting magnets. However, this is not an application of the Meissner effect; rather, it is just another application of the superconductor's zero resistance, and there are magnetically levitated trains that operate with ordinary (non-superconducting) magnets.

14.9 The Digital Information Age*

* Every student of science should have some appreciation of the digital revolution that is occurring now. Nevertheless, the material in the section will not be used again in this text and may be skipped without loss of continuity.

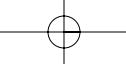
We conclude this chapter with a topic not normally included in modern physics books: the storage and transmission of digital information. One could argue that this topic is rightly the domain of computer science or electrical engineering, rather than physics. However, the technology produced by modern physics has lead directly to the rise of the digital information age we now inhabit. The digital revolution has radically altered the working life of every physicist, whether experimentalist or theorist, and every physics student should have some understanding of the science behind this revolution.

Right now, in the early twenty-first century, we are living in a time of transition from analog to digital technologies. In **analog** instruments, information is generated and transmitted as some kind of continuously varying signal, such as a voltage, a current, or some mechanical position. For instance, a galvanometer is an analog device, in which the magnitude of a current is measured by the position of a needle on a dial. Old vinyl records are analog devices in which the amplitude and frequency of sound waves are encoded in the continuously varying shape of a groove in a plastic disk. Prior to about 1950, virtually every electronic device and every storage medium in the world was analog. Then came computers, the first of a wave of digital devices. In a **digital** instrument, information is stored and transmitted as a discrete series of numbers, usually zeros and ones. In the 1980s, compact disks (CDs) replaced vinyl records as the storage medium of sound. In a CD the music is stored at a series of numbers, represented by long and short pits etched in a piece of plastic (see Fig. 14.41). Today, there are only a few analog technologies still in widespread use, but these are rapidly disappearing. DVDs have largely replaced VHS tapes. Digital cameras are quickly replacing film cameras. Digital television transmission is beginning to replace analog TV signals. Telephone transmissions are mostly digital. Movie projectors will be digital in a few years, and bulky film cans will be a thing of the past.



FIGURE 14.41

Long and short pits on the surface of a CD (compact disk) encode digital information. The longest pits are about $2 \mu\text{m}$ long. This feature size is large compared to present state-of-art photolithographic capabilities.



486 Chapter 14 • Solids—Applications

What has caused this revolution? Why are we switching from analog to digital? The answer is that digital technologies have two enormous advantages over analog ones: **noise immunity** and **information compression**. We begin by describing noise immunity.

Analog signals can be easily corrupted by noise — noise being any unwanted, uncontrolled distortion of the signal. The degradation of an analog signal is directly proportional to the noise: Double the noise, and the signal degrades twice as much. So with analog signals, you must constantly battle the noise — if even a little sneaks in, the signal corrupts. But digital signals are immune to moderate amounts of noise. So long as the noise is not so large as to blur the difference between a zero and a one, the signal is uncorrupted. With an analog signal, as noise gets worse, the signal degrades steadily, but with a digital signal, as the noise increases, the signal remains virtually perfect until the noise is large and then, rather suddenly, the digital signal is obliterated. With digital, it is usually all or nothing (see Fig. 14.42).

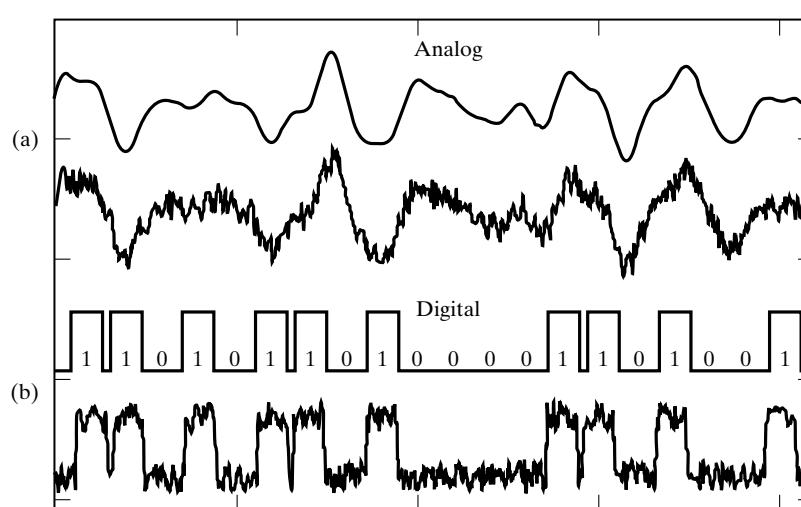
The noise immunity of digital data is even better than we have suggested. Even in the case where the noise is so large that a zero is occasionally mistaken for a one, **error-correcting codes** can detect mistakes and correct them. Whenever large amounts of digital data are transmitted over a communications channel (such as a television cable, a fiber optics phone line, or the read head on a hard disk), some additional, redundant data is sent along with the original signal. The redundant data acts as a check on the validity of the signal and can also provide enough information to reconstruct the signal if a not too extensive mistake is made. Here is a simple-minded example of an error-correcting code: If one wanted to send a stream of bits, such as

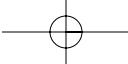
$$1 \ 0 \ 1 \ 1 \ 0 \dots$$

one could agree to send each bit three times:

$$111 \ 000 \ 111 \ 111 \ 000 \dots$$

so that if a single-bit error occurred during transmission, producing the message

$$111 \ 000 \ 110 \ 111 \ 000 \dots$$


Section 14.9 • The Digital Information Age **487**

then one could both detect and correct the mistake. This, however, is an extremely inefficient error-correcting code since it requires that the signal be tripled in length.

Much more efficient error-correcting codes exist, and these are based on a remarkable theorem discovered in 1949 by the Bell Labs mathematician and engineer Claude Shannon. **Shannon's theorem** states that it is always possible to design efficient error-correcting codes that can reconstruct a corrupted signal. Given a signal with a known error rate (say, a random error every 100 bits) Shannon's theorem allows one to compute the minimum number of bits that must be added to the original signal in order that it contain enough redundant information so that virtually all errors can be detected and corrected. The increased signal length due to the error-correcting code turns out to be modest. For instance, if the error rate is 1 bit in 100, then the signal length must be increased only 8%. Interestingly, Shannon didn't actually create these efficient error-correcting codes; rather, he proved that they existed, which provided the motivation for other mathematicians to find them. An example of an efficient error-correcting code is explored in Problem 14.49. Error-correcting codes are essential for reading DVDs covered with fingerprints, for transmitting files over the Internet, and for communicating with distant interplanetary probes (see Problem 14.51).

Besides noise immunity, the other great advantage of digital data over analog is the capability of **data compression**, which means using fewer bits to store the same amount of data. Compression schemes take advantage of the fact that signals almost always contain a high degree of repeated information. For instance, the English language has a high degree of redundancy. Consider the following sentence, which is missing 1/3 of its letters, yet is still readable:

Th qick brwn fx jmpd ovr th lzy dg.

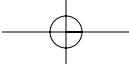
Such a signal can be rewritten so as to reduce the degree of redundancy, without losing essential information. Again, the work of Claude Shannon illuminated the problem. Shannon showed exactly how to compute the degree of redundancy in a signal and therefore how much the signal can be compressed without destroying any of its content. Most compression schemes work by searching the original signal for repeated sequences of bits (strings) and preparing a list of the most commonly occurring repeats. Then a dictionary is prepared that translates the longer, commonly occurring strings into shorter strings. Compression schemes can be **lossless**, meaning that the original signal can be restored exactly. Or they can be **lossey**, meaning that some loss of information is tolerated in return for a greater degree of compression. For instance, images are sent over the Internet in lossey compressed formats (jpegs and gifs), while “zipped” files are sent with lossless compression. DVD video is a lossey compression format that compresses the original video data by a factor of about 40. Analog television signals are allotted a 6 MHz bandwidth per channel. Digital TV (HDTV) will produce higher resolution with the *same* bandwidth through the signal compression.

As the microelectronics industry produces smaller and smaller devices, the density of information in digital storage media (hard disks, DVDs, etc.) is increasing rapidly. Using the present deep-UV lithographic technology, each bit takes up an area of at least $(0.1 \mu\text{m})^2 = 10^{-14} \text{ m}^2$, which limits storage density to about 10^{10} bits per cm^2 . However, with the new scanning probe technologies (Section 14.7), we have begun to develop the ability to manipulate

Claude Shannon
(1916–2001, American)



Trained first as an electrical engineer, then as a mathematician, Shannon made fundamental contributions to the theory of digital computation while still a student. In 1948, while working at Bell Laboratories, he published “A Mathematical Theory of Communication,” a paper that set the stage for the digital age. Known for his ability to approach difficult problems from new directions, he was often seen in the hallways of Bell labs late at night, riding a unicycle while juggling. He was also a pioneer in the field of artificial intelligence; the photo shows him working on an electromechanical mouse that learns to find its way through a maze.



488 Chapter 14 • Solids—Applications

FIGURE 14.43

The letters IBM written with individual xenon atoms on the surface of nickel. The atoms were manipulated and imaged with an STM. Each letter is about 1.2 nm high. In order for the letters to remain stable, the sample temperature must be held at 4 K, so that the xenon atoms do not diffuse because of thermal motion.



individual atoms. If atoms can be placed in a controlled way on a surface or (even better) throughout a volume, the information density that can be achieved is truly astounding (see Figs. 14.31 and 14.43 and Problem 14.50).

Example 14.3

Information content of the human race. In Harlan Ellison's science fiction story "Demon with a Glass Hand," dramatized in a 1964 episode of *Outer Limits*, the human race escapes a radioactive plague by being safely stored as digital information, which will be used to reconstruct humanity when the plague recedes. The future earth's entire population of 10 billion people is encoded into the atomic structure of a copper–tungsten alloy wire, which is stored inside the hero's body (Fig. 14.44). The human genome (our complete genetic code) contains about 10^{10} bits of information. Recent studies indicate that the human brain can store no more than about 10^{12} bits of information in its neural interconnects (and probably much less than this). Is it possible to store 10 billion people in a small piece of metal?

If 0s and 1s are stored as Cu or W atoms in an alloy, the volume of a single atom is required to store one bit of information. The volume per atom in a solid is about $(0.3 \text{ nm})^3 \approx 3 \times 10^{-29} \text{ m}^3$. Assuming 10^{12} bits are required to store one human's total information (a lifetime of memories plus a little extra for the DNA in *uncompressed* format), the information content of the entire population of 10 billion people is no more than $10^{12} \times 10^{10} = 10^{22}$ bits. Notice that this is less than Avogadro's number. The volume of Cu–W alloy required to store this information is therefore

$$10^{22} \text{ bits} \times (3 \times 10^{-29} \text{ m}^3/\text{bit}) = 3 \times 10^{-7} \text{ m}^3 \approx (0.007 \text{ m})^3$$

FIGURE 14.44

In the 1964 *Outer Limits* episode, "Demon with a Glass Hand," the protagonist is informed by his talking electronic hand that he is the guardian of an important storage medium, a small Cu–W alloy wire. Encoded in the arrangement of atoms in the wire is the information needed to reconstruct the entire human race.



This information can be stored in a cube about 7 mm on a side. With data compression (and especially considering that all humans are very similar), the information could be stored in a much smaller volume. As Richard Feynman once said, “There’s plenty of room down there.”

CHECKLIST FOR CHAPTER 14

CONCEPT	DETAILS
Semiconductor Holes Donor and acceptor impurities	Insulator with small band gap, $E_g \approx 1$ eV (Sec. 14.2) Unoccupied states in full valence band = mobile, positive charges At room temperature, each donor creates a conduction electron, each acceptor creates a hole
<i>pn</i> Junction diode	Low resistance in forward bias, high resistance in reverse bias, $I(V) = I_R(e^{eV/kT} - 1)$ (14.4)
Transistor Bipolar junction transistor (BJT) Field-effect transistor (FET)	Solid-state amplifier or switch (Sec. 14.4) Emitter, base, and collector: $I_c = \beta I_b$ Source-drain resistance controlled by gate voltage
LEDs, solar cells, CCD arrays	Semiconductor applications (Sec. 14.5)
Integrated circuits	Microelectronic device with large number of transistors fabricated by photolithography (Sec. 14.6)
Scanning tunneling microscope Work function ϕ Tunneling depth λ	Capable of atomic resolution (Sec. 14.7) Binding energy of conduction electron in metal, $\phi \approx 4$ eV Wave function leaks into forbidden region, $\psi(x) = \psi_0 e^{-x/\lambda}$
Superconductivity Meissner effect BCS theory	Zero resistivity below the critical temperature T_c (Sec. 14.8) $B = 0$ in a superconductor Electron–phonon interaction forms Cooper pairs, Bose–Einstein condensation
Digital technology Noise immunity Data compression	Signal storage and transmission as series of numbers (Sec. 14.9) Shannon’s theorem, error-correcting codes Removing redundant information reduces message length

PROBLEMS FOR CHAPTER 14

SECTION 14.2 Semiconductors

- 14.1** • Consider a single crystal of an insulator with a band gap of 3.5 eV. What colors of visible light will this crystal transmit?
- 14.2** • Crystalline sulfur absorbs blue light but no other colors; as a result, it is a pale yellow transparent solid. It is also an electrical insulator. From this information, deduce the width of the band gap of crystalline sulfur, and describe the occupancy of the bands above and below the gap.
- 14.3** • Germanium is a semiconductor with a band gap of 0.7 eV. What colors of visible light does it transmit?
- 14.4** • Explain why many insulators are transparent to visible light, whereas most semiconductors are opaque to visible light but transparent to infrared. (Some far-infrared cameras use lenses made of germanium, which block all visible light but pass and focus the IR light.)
- 14.5** • The longest wavelength of electromagnetic radiation that is absorbed in gallium arsenide is 890 nm. What is the band gap of this semiconductor?
- 14.6** • What is the longest wavelength radiation that can produce a current in a photovoltaic cell that consists of a silicon *pn* diode with band gap of 1.1 eV?
- 14.7** • (a) If a crystal of pure silicon is doped with atoms of phosphorus, what type of semiconductor results, *n* or *p*? (b) What if the same crystal were doped with aluminum? (c) With antimony?
- 14.8** • (a) If a crystal of pure germanium is doped with atoms of boron, what type of semiconductor results, *n* or *p*? (b) What if the same crystal were doped with indium? (c) With arsenic?
- 14.9** •• A single crystal of silicon is doped with both donor and acceptor impurities, and the temperature is high enough that all impurity sites are ionized. The concentration of donor impurities is twice that of the

490 Chapter 14 • Solids—Applications

acceptor impurities. Is the net charge of the crystal positive, negative, or zero? Explain.

- 14.10 •• (a)** What is the concentration (number per volume) of atoms in a single crystal of silicon? Silicon has a density of 2.33 g/cm^3 . **(b)** Qualitatively, how does this compare to the concentration of thermally activated conduction electrons in pure silicon at room temperature, $n_i \approx 10^{10} \text{ cm}^{-3}$?

- 14.11 ••** What is the concentration of conduction electrons in an n -type sample of silicon if one out of every million silicon atoms is an arsenic atom? Assume that the temperature is high enough that all the donor atoms are ionized. The density of silicon is 2.33 g/cm^3 .

- 14.12 ••** A typical dopant concentration in a silicon device is about 10^{22} m^{-3} . **(a)** What is the ratio of the number of silicon atoms to the number of dopant atoms at this concentration? The density of silicon is 2.33 g/cm^3 . **(b)** If a manufacturing process requires that there be less than 1 uncontrolled impurity atom for every 1000 (controlled) dopant atoms, what is the allowable impurity concentration in parts per billion (1 ppb = 1 impurity atom per billion silicon atoms)?

- 14.13 ••** As a nuclear particle traverses a particle detector, it creates electron-hole pairs and loses energy. The average energy lost per pair created is 3.1 eV. (This is much bigger than the band gap of 1.1 eV because many of the excited electrons come from deep down in the valence band, and other processes — like inelastic scattering — take some energy.) A particle with initial kinetic energy of 2 MeV enters the detector and is brought to rest. **(a)** How many electron-hole pairs does it create? **(b)** What total charge flows around the circuit?

- 14.14 •• (a)** Use equation (14.2) to make a semilog plot of conduction electron concentration versus $(1/kT)$ for intrinsic (pure) silicon. The conduction electron concentration in Si at $T = 293 \text{ K}$ (room temperature) is 10^{10} cm^{-3} . **(b)** Below what temperature are there essentially no charge carriers in pure Si (say, less than one carrier per cm^3)? **(c)** Is there some temperature below the melting point of Si ($T_{\text{melting}} = 1690 \text{ K}$) at which the conduction electron concentration in Si is comparable to that of a metal?

- 14.15 •• (a)** Use equation (14.2) to compute the ratio of the conduction electron concentrations in Si and Ge at room temperature (293 K) and **(b)** in Si and GaAs at room temperature. [Assume that the prefactor in (14.2) is approximately the same for all these materials.]

- 14.16 ••** The donor state wave functions in silicon have small binding energies ($E_{\text{binding}} \approx 0.05 \text{ eV}$) and large Bohr radii ($a_d \approx 3 \text{ nm}$) for two reasons: (1) Because of the interaction between the lattice and the “nearly free” electrons in the conduction band, the electrons behave as if their mass was *smaller* than the free-electron mass. The “effective mass” in silicon is $m^* \approx 0.2 m_e$. (2) The polarization of the silicon lattice partially shields the $+1e$ charge of the impurity site. In general, the force of electrostatic attraction between two charges in a polarizable medium is given

by $F = kq_1q_2/(Kr^2)$, where K is a dimensionless number called the *dielectric constant*. In silicon, $K = 11.7$. **(a)** Derive expressions for the Bohr radius and the ionization energy of the ground state of a donor wave function with effective mass m^* and in a medium with dielectric constant K . **(b)** Compute the values of the ionization energy and Bohr radius of a donor state in silicon. (Compare your values with the known values for phosphorus in silicon of $E_{\text{binding}} = 45 \text{ meV}$ and $a_{\text{donor}} = 3.0 \text{ nm}$.)

- 14.17 ••** In direct-gap semiconductors, such as GaAs, electrons near the bottom of the conduction band and valence-hole states near the top of the valence band have momenta near zero. In indirect-gap semiconductors, such as Si and Ge, these electron and hole states have very different momenta. **(a)** Compute the momentum of a visible-wavelength photon. (Recall that for photons, $p = E/c$.) **(b)** Compute the momentum of an electron with a wavelength of 0.3 nm, which is typical for a conduction electron in a semiconductor. **(c)** Explain why conservation of momentum forbids production of photons from electron-hole recombination in silicon.

- 14.18 ••• The Hall effect and the sign of charge carriers.** Consider a rectangular conductor of width w and thickness t carrying a current I in a uniform magnetic field \mathbf{B} perpendicular to the plane of the conductor, as shown in Fig. 14.45. The charge carriers, with charge q and drift velocity \mathbf{v} , experience a Lorentz force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$, which tends to push them to one side of the conductor. The resulting steady-state charge buildup causes a voltage difference V_H , called the *Hall voltage*, between the sides of the conductor. **(a)** Show that the Hall voltage is given by

$$V_H = \frac{BI}{nqt} \quad (14.14)$$

where n is the concentration of charge carriers. [Hint: In steady state, there is no net force on the charge carriers transverse to the current direction, so the force due to the transverse E field must be canceled by the force from the B field.] **(b)** Show, with an appropriate diagram, that the sign of the Hall voltage gives the sign of the charge carriers. Thus, the Hall voltage allows one to determine whether the carriers in a doped semiconductor are electrons or holes. **(c)** Compute the magnitude of the Hall voltage in an n -type semiconductor sample at room temperature with donor concentration

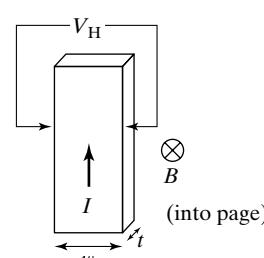


FIGURE 14.45

(Problem 14.18.) The Hall effect.

$N_d = 10^{22} \text{ m}^{-3}$ and the following parameters: $t = 1 \text{ mm}$, $I = 0.01 \text{ A}$, $B = 1.0 \text{ T}$.

- 14.19 ••** In an intrinsic semiconductor, the concentration of conduction electrons equals the concentration of valence holes: $n = p = n_i$, where n is a temperature-dependent parameter called the **intrinsic concentration**. In a doped semiconductor, the product $np = n_i^2$ turns out to be *independent* of the donor concentration N_d or acceptor concentration N_a . For instance, as donors are added, n increases, but p decreases (due to increased rate of electron-hole recombination) and np remains fixed. The fact that the product np is independent of impurity concentration is called the **law of mass action**. **(a)** Write a formula for the hole concentration p for a donor-doped semiconductor, in terms of n_i and N_d , assuming that $N_d \gg n_i$ and the temperature is high enough that all donors are ionized. **(b)** In silicon at room temperature, $n_i = 4.6 \times 10^{10} \text{ cm}^{-3}$. What donor concentration will result in a hole concentration of $10^{-4} n_i$?

- 14.20 ••** *The metal-insulator transition in silicon.* If the concentration of donor impurities in silicon is large enough for the donor state wave functions to overlap, the donor states will hybridize to form a partially filled band of delocalized wave functions, and the heavily doped silicon will become a metal (that is, it will conduct at $T = 0 \text{ K}$). The Bohr radius of a donor impurity is about 3 nm in silicon and about 8 nm in germanium. Estimate the donor concentrations at which silicon and germanium undergo the metal-insulator transition.

SECTION 14.3 The *pn* Junction Diode

- 14.21 •** The diode circuit shown in Fig. 14.46 is called a half-wave rectifier. The input to this circuit is a sinusoidal voltage $V(t) = V_0 \sin \omega t$. Sketch the output voltage. Assume that the diode has infinite resistance in reverse bias and zero resistance in forward bias.

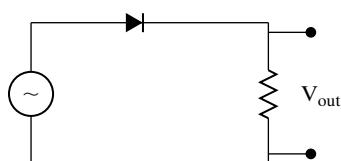


FIGURE 14.46

(Problem 14.21.) Half-wave diode rectifier.

- 14.22 ••** The arrangement of diodes shown in Fig. 14.47 is called a full-wave rectifier. The input to the circuit is a sinusoidal voltage $V(t) = V_0 \sin \omega t$. Sketch the output

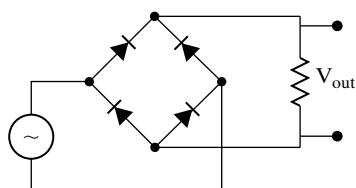


FIGURE 14.47

(Problem 14.22.) Full-wave diode rectifier.

voltage. Assume that the diode has infinite resistance in reverse bias and zero resistance in forward bias. (The full-wave rectifier is the first stage in a power supply circuit that converts AC voltage into DC voltage.)

- 14.23 •• (a)** Consider the diode equation (14.4) and show that for large negative voltages, the current is $I = -I_R$. (This means there is a current in the reverse direction, but it is very small, as you will now show.) **(b)** Consider forward and reverse voltages of 1 volt ($V = \pm 0.5$ volt). Taking $T = 293 \text{ K}$ (room temperature), write down the corresponding forward and reverse currents and compute their ratio.

- 14.24 •• (a)** Is the ordinary current in a copper wire a drift current or a diffusion current? **(b)** Give an example of a diffusion current.

- 14.25 •• Electric field in the depletion region.** Consider a silicon *pn* junction diode, with donor concentration N_d on the *n* side, and acceptor concentration N_a on the *p* side. In equilibrium, the uncompensated impurities on either side of the junction form two slabs of equal and opposite charge density as shown in Fig. 14.48. The potential difference between the two sides of the depletion region is the contact potential $V_0 = 0.7 \text{ V}$, and the total width of the depletion region is $2L$. **(a)** Use Gauss's law to derive an expression for the electric field E as a function of position within the depletion region. Sketch the electric field versus position. [Hint: Apply Gauss's law to the cylinder shown in the figure.] **(b)** Derive a formula for the half-width L of the depletion region in terms of the contact potential V_0 . [Hint: Recall that $\Delta V = \int E dx$.] **(c)** Assuming $N_d = N_a = 10^{22} \text{ m}^{-3}$ (a typical impurity concentration), what is the width of the depletion region and what is the magnitude of the electric field at the junction? The dielectric constant of silicon is $K = 11.7$.

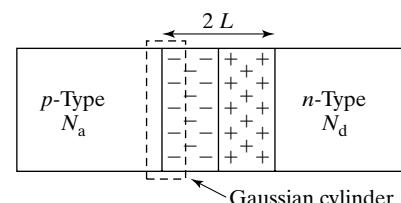


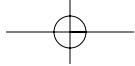
FIGURE 14.48

(Problem 14.25.) Depletion region in a *pn* junction.

SECTIONS 14.4 and 14.5 (The Transistor, Further Semiconductor Applications)

- 14.26 •** In a transistor the ratio $I_c/I_b = \beta \approx 100$. What is the ratio I_e/I_b ?

- 14.27 •** Estimate the energy output of a solar cell on a sunny day. When the sun is overhead, the power incident on each square meter of the ground is about 1000 W. Assuming a solar-cell efficiency of 15%, estimate the daily energy output of a solar panel with an area of 10 square meters. Give the answer in joules and in kilowatt-hours.



492 Chapter 14 • Solids—Applications

14.28 •• Consider a current of 3 mA flowing through an LED whose band gap is about 2 eV. **(a)** How many photons are produced per second? **(b)** What is the power of the light produced? Notice that this small current produces a light that is easily visible. (For comparison, a typical flashlight bulb, which is very easily visible, consumes about a watt of power, of which about 50 mW is visible light). **(c)** Show in general that this power (in watts) is numerically just the product of the current (in amperes) and the band gap (in eV).

14.29 ••• Photon counting in CCD cameras. A typical CCD camera has pixels with dimensions $5 \mu\text{m} \times 8 \mu\text{m}$ and a lens 2 cm in diameter. Suppose the lens focuses light from a distant point source onto a single pixel. **(a)** How many visible photons fall on the CCD pixel that is exposed to the light from a 100-W lightbulb 1 mile away for 1 s. **(b)** What is the charge produced in the CCD pixel by this light?

SECTION 14.6 Integrated Circuits

14.30 • (a) What is the smallest feature size that can be produced with photolithography using visible light? Which is capable of producing smaller features: visible light, UV light, or X-rays?

14.31 • In electron-beam lithography, a focused beam of electrons (produced by a scanning electron microscope) is used instead of UV light to expose photoresist. Which is capable of producing smaller features: 100 eV electrons or 50 keV electrons? (In practice, the resolution of electron-beam lithography is limited by aberrations in the electron optics of the electron microscope, rather than by the wavelength of the electron — the aberration-limited resolution is about 5 nm.)

14.32 •• “Thermal management” is the electronics industry’s term for “don’t fry the chip.” The Pentium 4 chip sits inside an epoxy package with 478 pins. The CPU itself has an area of 217 mm^2 , contains 42 million transistors, and dissipates 70 W of power when running at full throttle. Measure the surface area of a 100-W lightbulb, and compare the power dissipated per area of a 100-W lightbulb to that of a Pentium 4 chip. Pentium chips have a “thermal throttle” that automatically reduces power dissipation (and performance) of the chip if it overheats.

14.33 •• The ultimate limit of Moore’s law. In the year 2001 the line width in Pentium chips was $0.18 \mu\text{m}$. For the past 3 decades, the line width in integrated circuits has decreased an average of 11% per year. Assume that this trend will continue, and make a graph of line width versus year. If this trend continues, in what year will the size of electronic devices decrease to the size of atoms? (It seems obvious that electronics cannot get smaller than single atoms, so Moore’s law must fail at that level, if not before.)

SECTION 14.7 The Scanning Tunneling Microscope

14.34 • Make a semilog plot showing tunneling current versus tip-sample separation for an STM. Assume that the tunneling depth is $\lambda = 0.1 \text{ nm}$ and the tunneling

current is 1 nA when the separation is $z = 0.2 \text{ nm}$. In an STM the circuitry usually limits the measurable currents to the range between $100 \mu\text{A}$ and 10^{-14} A .

14.35 • Which of the metals listed in Table 14.2 will produce the largest thermionic emission current? Explain your reasoning.

14.36 •• In the forbidden region the wave-function amplitude decays as (14.11) $\psi(x) = \psi_0 e^{-x/\lambda}$. Define P as the ratio of the probabilities of finding the electron at x and at 0: $P = |\psi(x)/\psi_0|^2$. **(a)** Derive a formula for x in terms of P . **(b)** For a tunneling depth of $\lambda = 0.1 \text{ nm}$ (typical of wave-function decay outside a metal), what is the distance x at which $P = 10^{-10}$? 10^{-20} ?

14.37 •• Show how the electric field between the two metals in Fig. 14.27(b) depends on the slope of the electron energy dU/dx .

14.38 •• Argue that a tunnel junction between two metals has an *ohmic* resistance; that is, the current I is proportional to the applied voltage V . [Hint: Consider Fig. 14.27(b), and argue that the current is proportional to the number of occupied electronic states on the left that can tunnel into empty states on the right. Note that the number of states per energy interval dN/dE (called the density of states) is very nearly constant for energies within an eV of the Fermi level.]

14.39 •• Assume that a significant field-emission current occurs when the barrier width is smaller than some critical distance d , as shown in Fig. 14.49. **(a)** Derive an expression for the minimum electric field needed to produce field emission. **(b)** Assuming that the critical distance for tunneling is $d = 2.5 \text{ nm}$ and the work function is $\phi = 4 \text{ eV}$, what is the electric field needed for field emission?

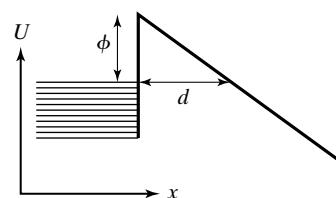


FIGURE 14.49

(Problem 14.39.)

14.40 ••• Derivation of the field-emission equation. Show that the tunneling current in field emission is given by $I \propto \exp(-4\sqrt{2\pi}\phi^{3/2}/3\hbar eE)$, where E is the electric field. This is a difficult problem because the tunneling barrier does not have constant height. See Section 17.10 for direction in how to deal with this kind of tunneling problem.

SECTION 14.8 Superconductivity

14.41 • Use equation (14.13) to compute the gap energy E_g of lead ($T_c = 7.2 \text{ K}$).

14.42 •• (a) Use Faraday's law to argue that a perfect conductor (one with zero resistivity) must have a constant B field in its interior. **(b)** What is the difference between a superconductor and a perfect conductor?

14.43 •• (a) Use Ampere's law to show that the magnetic field in a wire of radius r carrying a current I has a maximum value at the surface of the wire given by $B = \mu_0 I / (2\pi r)$. **(b)** Compute the maximum current in a Pb wire 1 mm in diameter at 4.2 K. At 4.2 K the critical field of Pb is 0.06 T.

14.44 •• The temperature dependence of the critical field B_c has the functional form

$$B_c(T) = B_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \quad (14.15)$$

This formula describes either B_{c1} or B_{c2} in type II superconductors. Values of $B_c(0)$ are given in Table 14.3. **(a)** What is the critical field for niobium at 4.2 K? **(b)** A type II superconductor with a critical temperature of $T_c = 23$ K is found to have a critical field of 25 T at 4.2 K. This material is used to make a $B = 10$ T magnet. What is the maximum temperature at which this 10 T magnet can operate?

14.45 ••• Use the Heisenberg uncertainty principle to estimate the size of a Cooper pair in indium, where the gap energy is 0.0011 eV and the Fermi energy is 8.6 eV. (Recall that the gap energy is the binding energy of a Cooper pair.) The Cooper pair, consisting of two conduction electrons, has an energy of twice the Fermi energy reduced by the gap energy. If the Cooper pair is squeezed down from a very large size to a size Δx , the momentum must increase by an amount Δp given by the uncertainty principle, and consequently the energy will increase by $\Delta E = (dE/dp)\Delta p$. The energy increase cannot exceed the binding energy, or else the pair will unbind. These considerations allow you to compute the minimum possible size of a Cooper pair.

SECTION 14.9 The Digital Information Age

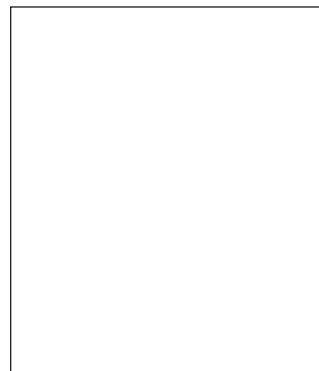
14.46 •• Compute the number of bits required to store an *uncompressed* image. **(a)** Suppose one wanted to have 128 shades of gray (ranging from pure white to pure black) in the image, how many bits per pixel would be required to specify the gray level or brightness? **(b)** Consider a color image of 640×480 pixels (standard VGA resolution), in which each pixel has separate red, green, and blue levels of brightness. How many bits are required to store this image with no compression? Compare this with the size of a typical VGA jpeg image stored on your PC. (Don't forget that on a PC, file sizes are given in bytes and 1 byte = 8 bits.)

14.47 •• Examine Fig. 14.41, and estimate the number of pits on a CD.

14.48 •• The Library of Congress, which is the United States' official repository of all printed documents, contains approximately 120 million books. **(a)** Estimate the

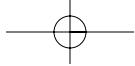
number of characters (letters, numbers, and spaces) in all the books in the Library of Congress. **(b)** Estimate the number of bits (1s and 0s) of information required to store the library in uncompressed format. **(c)** If these bits were stored on the surface of a copper-tin alloy as individual atoms (with a Cu atom meaning 1, and an Sn atom meaning 0, say), what would be the area required to store the whole library? **(d)** Using Shannon's information theory, it can be shown that ordinary English text can be stored in a compressed format that contains only about 2 bits per letter. What surface area would be required to store the entire library in this compressed format? **(e)** If this compressed information was stored in the interior of a cube of metal alloy, how large would the cube be?

14.49 •• An efficient error-correcting code. Consider a signal consisting of 16-bit message blocks accompanied by 8-bit check blocks; that is, $1/3$ of the signal consists of check bits. Each 16-bit message block is arranged into a 4×4 array, and the 8 check bits are aligned with the top row and left column of the message, as shown, with the 16-bit message in the dark square in the lower right.



Each check bit is set so that there are an even number of 1s in every column and every row (in those 4 rows and 4 columns containing message bits). You should verify that this is true for the (message + check) array shown here. **(a)** Suppose a single-bit error occurs somewhere in the message block. How many check bits will violate the "even sum" rule? How can this information be used to identify and correct the mistake? **(b)** Suppose a single-bit error occurs somewhere in the check block. How many check bits will violate the "even-sum" rule? How does this information indicate that the error is not in the message, but in the check block? **(c)** Suppose that $N \times N$ message blocks are used instead of 4×4 message blocks. How many check bits are required? What fraction of signal bits are check bits? **(d)** Explain why this simple scheme fails if two errors occur in the same row or column.

14.50 •• (a) If information can be stored in the placement of individual atoms of a solid, what information density (in bits per cm^3) is possible? Recall that in most



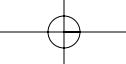
494 Chapter 14 • Solids—Applications

solids, each atom takes up a volume of about $(0.3 \text{ nm})^3$. **(b)** How large a volume is required to store 1 gigabit? 1 terabit? 1 gigaterabit = 10^{21} bits?

14.51 ••• Communications with interplanetary space probes.

When an incoming digital signal has very low power, the message must be sent very slowly, as this problem will show. **(a)** Consider a space probe, a great distance R from the earth, transmitting a radio signal at wavelength λ and power P_0 from a radio dish of diameter d . On earth is a very large radio receiver with a dish of diameter D , collecting the radio-wavelength photons sent by the distant probe. Suppose the 1s and 0s of the digital message are represented by radio signals of high and low intensity, with a ratio of 4:1. Also suppose that in order to distinguish the signal from background noise, the receiver must collect at least 10^4 photons per “one bit” (and 2500 photons per “zero bit”). Derive an approximate formula for the highest bit rate (number of bits per second) at

which the space probe can transmit a message which can be read by the receiver on earth. [Hint: Don’t forget diffraction effects: The outgoing radio signal has an angular divergence of $\Delta\theta \approx \lambda/d$.] **(b)** In 1995 NASA’s Galileo spacecraft entered orbit around the planet Jupiter, 800 million kilometers from earth. Galileo has a main radio antenna 5 meters in diameter that was designed to transmit power $P_0 = 20 \text{ W}$ at wavelength $\lambda = 3 \text{ cm}$. NASA’s radio receiver on earth has a diameter of 70 meters (big!). Galileo’s main antenna was designed to transmit information at what bit rate? **(c)** Unfortunately, the main antenna on Galileo failed early in the mission. A small backup antenna operating at the same power and wavelength as the main antenna, but with an effective diameter of only 5 cm, was used to transmit data to earth. At what bit rate can this backup antenna send data? A typical image from Galileo’s camera contains about 10^6 bits. How long does it take for a picture to be sent from Galileo?



Chapter 15

Statistical Mechanics

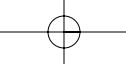
- 15.1 Introduction
 - 15.2 Temperature
 - 15.3 The Boltzmann Factor
 - 15.4 Counting Microstates: The Equal-Probability Hypothesis
 - 15.5 The Origin of the Boltzmann Relation*
 - 15.6 Entropy and the Second Law of Thermodynamics*
 - 15.7 The Quantum Ideal Gas — a Many-Particle System
 - 15.8 Energy and Speed Distributions in an Ideal Gas
 - 15.9 Heat Capacities
- Problems for Chapter 15

* These sections can be skipped without loss of continuity.

15.1 Introduction

Statistical mechanics is the mechanics of systems with very large numbers of constituent particles — a liter of air with some 10^{22} molecules of N₂ and O₂, a cup of water with some 10^{25} water molecules, a meter of copper wire with some 10^{23} conduction electrons. There are classical and quantum versions of statistical mechanics. While both have applications under the heading of “modern physics,” it is the quantum version that is the more important and, rather surprisingly, the simpler. This chapter is a brief introduction to statistical mechanics, with an emphasis on the quantum version.

The name “statistical mechanics” reflects that because of the huge number of particles involved, we can treat them *statistically*, ignoring the detailed motion of the individual particles. This has the surprising consequence that problems in statistical mechanics (either classical or quantum) are often actually easier than problems involving only three or four particles whose motions need to be followed in detail. The detailed specification of the microscopic state, or **microstate**, of any system of N particles involves giving the wave functions of all N particles (in the quantum case) or the positions and velocities of all N particles (in the classical case). When N is large, this represents an appalling amount of information; to write down the microstate of a mole of gas would require more than the world’s total supply of paper! Fortunately, in statistical mechanics, we have no desire to know the microstate of a system. Rather, we want to know just a small number of *macroscopic* variables that are statistical averages over the many microscopic variables. For example, to specify the macroscopic state, or **macrostate**, of a container of helium gas, we have only to give the volume V , the number of atoms N , and the total energy E (or, equivalently — as we will see — the temperature T). Statistical mechanics is the study of the relation between the observable macrostates and the underlying, but unobservable and ultimately uninteresting, microstates.



496 Chapter 15 • Statistical Mechanics

15.2 Temperature

We begin by discussing the meaning of temperature, a key idea in statistical mechanics. What is temperature? Before the development of statistical mechanics and kinetic theory (Section 3.7), temperature was defined in a purely pragmatic manner as “the thing that thermometers measure.” But this definition shed no light on the physical microscopic meaning of temperature. Before kinetic theory was established, physicists could say only that hot things have higher temperature, cold things have lower temperature, and one can define a temperature scale that everyone agrees on, so that different thermometers give consistent results. But what, microscopically, is the difference between hot objects and cold objects? What happens to water as it is heated? The appearance and mass of the water remain unchanged. What property is your body sensing when you dip your hand in water and feel hot or cold? It was a mystery.

Today, we understand that temperature is a measure of thermal energy — the random, microscopic energy of atoms. You can think of temperature as a measure of the “jiggling” of atoms. At $T = 0$ K, the atoms of a material have no energy, or more precisely, they have as little energy as possible, with the system residing in its quantum ground state. A classical description of a solid material at $T = 0$ is an assembly of stationary atoms, rigidly fixed in position. As the temperature rises, the atoms begin to vibrate about their equilibrium positions, gently jostling their neighbors. As the temperature rises further, the atoms jiggle more energetically, crashing against their neighbors until the chemical bonds between atoms begin to break and the atoms slide past each other — the phenomenon of melting. With the proper conditions of high temperature and low pressure, the atoms enter a gaseous state, with weakly interacting atoms flying in straight lines between collisions. In this state, higher temperature means more kinetic energy of the flying atoms. When the hand is dipped in hot water the sensation of heat is caused by jiggling water molecules pounding against the atoms of delicate skin cells. The atoms of the cells vibrate in response, sending complex chemical messages along our nerves.

To quantify this notion of jiggling atoms, this microscopic energy, and to give a precise definition of temperature, we must first introduce the concept of **degrees of freedom**. A degree of freedom is a kind of microscopic repository of energy, a mode in which an atom can store energy. In the context of statistical mechanics, the number of degrees of freedom is defined as the number of squared terms in the detailed expression for the total energy. For example, a single atom in an ideal gas has only kinetic energy (no potential energy) and it can move in any of three orthogonal directions x , y , and z , so its energy is $E = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2$. There are three quadratic terms in this expression, so we say that the atom has three degrees of freedom. An ideal gas of N atoms has $3N$ degrees of freedom — three degrees of freedom for each atom. A particle on a spring, moving in one dimension only, has both kinetic energy and potential energy (stored in the spring). The energy is $E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2$. Here there are two quadratic terms, and hence, two degrees of freedom. A solid consisting of N atoms held in place by springlike chemical bonds between nearest neighbors (Figure 15.1) has $6N$ degrees of freedom because the energy of each atom can be written as

$$E = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2 + \frac{1}{2}kx^2 + \frac{1}{2}ky^2 + \frac{1}{2}kz^2$$

where (x, y, z) is the displacement from the equilibrium position.

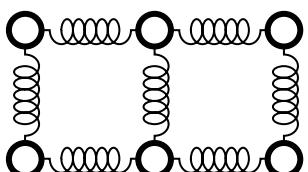


FIGURE 15.1

Ball-and-spring model of a solid.

We can now state an important theorem in statistical mechanics called the **equipartition theorem**. The equipartition theorem says that for a classical system in thermal equilibrium, on average, the total energy of the system is shared, or “partitioned,” *equally* among the various degrees of freedom. As an example, consider a gas of atoms. At any instant, the atoms possess a wide variety of velocities. Some atoms are fast, some are slow, a few are moving very nearly along the x -axis, others are moving in other directions. As the atoms collide, they exchange energy, and a given atom has a rapidly fluctuating energy and velocity as it zigzags from one collision to the next. However, in the long run, each degree of freedom of each atom, has exactly the same *average* energy. This situation can be likened to a box of marbles being violently shaken. All the marbles shake with the same average energy. This will be true even if the marbles have different masses. If one of the marbles in the box is unusually massive, it will move about with a smaller than average speed, but its average kinetic energy ($\frac{1}{2}mv^2$) will be the same as the other marbles. If we arrange things so that the marbles on the right half of the box are moving much faster than the marbles on the left half (if the total energy is not shared equally among the available degrees of freedom), the system is *not* in thermal equilibrium. But the random shaking of the box will quickly even out the energy distribution, and equilibrium will be achieved.

As another example of the equipartition theorem, consider a mass on a spring, constrained to move in one dimension. The energy has two terms $E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2$; there are two degrees of freedom. If the mass-spring system is shaken, the energy will sometimes be all kinetic (when $x = 0$) and sometimes all potential (when $v = 0$ and the mass is at the turning point). But on average, the energy is shared equally between the two terms.

Although we have not proved the equipartition theorem, we hope it seems plausible. A semirigorous proof is explored in Problem 15.9. There is one detail in the statement of the theorem that is easy to overlook, but will be important later. The equipartition theorem begins: “For a *classical* system in thermal equilibrium . . .” This caveat means that quantum effects are ignored. In particular, the equipartition theorem assumes that the total energy of the system is continuous — not quantized, not discrete. We will see later that the quantization of energy in quantum systems can lead to a breakdown of the classical equipartition theorem and an unequal sharing of energy among the degrees of freedom.

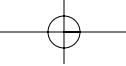
Now that we understand the concept of degrees of freedom and the equipartition theorem, we can give a precise definition of temperature. The temperature T of a classical system in thermal equilibrium is defined by the relation

$$\frac{1}{2}kT = (\text{average energy in each degree of freedom}) \quad (15.1)$$

where k is called Boltzmann’s constant* and has the experimentally determined value

$$k = 1.38 \times 10^{-23} \text{ J/K} = 8.62 \times 10^{-5} \text{ eV/K}$$

*Boltzmann’s constant k must not be confused with either the Coulomb constant $k = 1/(4\pi\epsilon_0)$ or the wave number $k = 2\pi/\lambda$. Almost invariably, Boltzmann’s constant appears next to the absolute temperature in the combination kT , so the notation is usually clear. When there is any danger of confusion, we will use the symbol k_B for Boltzmann’s constant.



498 Chapter 15 • Statistical Mechanics

The factor $\frac{1}{2}$ appears in (15.1) so that another equation (the Boltzmann relation, below) does *not* contain a factor of 2. The physically important concept here is that temperature is proportional to the average energy per degree of freedom, the amount of atomic “jiggling.”

Example 15.1

What is the average translational kinetic energy of a nitrogen molecule in air at room temperature, $T \approx 293$ K? What is the molecule’s rms speed?

The translational kinetic energy of a gas molecule is $E = \frac{1}{2}mv^2 = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2$, where v is the speed and v_x , v_y , and v_z are the components of the velocity. (In addition to translational kinetic energy, diatomic molecules, such as nitrogen or oxygen, can have rotational kinetic energy. See Section 15.9). There are three degrees of freedom in this expression, and each has an average energy of $\frac{1}{2}kT$. The total average kinetic energy is therefore

$$\frac{3}{2}kT = \frac{3}{2}(1.38 \times 10^{-23}) \times (293) \approx 6.1 \times 10^{-21} \text{ J} = 0.038 \text{ eV}$$

Notice that the answer does not depend on any of the properties of the molecule, such as mass. Any molecule in a gas at temperature T will have an average translational kinetic energy of $\frac{3}{2}kT$. In estimating thermal energies like this, it is worth remembering that the factor kT , which characterizes thermal atomic energy, has the approximate value 0.025 eV at room temperature.

The average kinetic energy is often written $\langle \frac{1}{2}mv^2 \rangle = \frac{1}{2}m\langle v^2 \rangle$, where the brackets $\langle \dots \rangle$ indicate an average over thermally fluctuating values. As discussed in Section 3.8, the root-mean-square (rms) value of the speed v is defined as $v_{\text{rms}} = \sqrt{\langle v^2 \rangle}$. Setting the average kinetic energy equal to $\frac{3}{2}kT$, we have

$$\frac{3}{2}kT = \frac{1}{2}m\langle v^2 \rangle$$

Solving for v_{rms} and recalling that the mass of an N_2 molecule is 28 atomic mass units, we get (using SI units)

$$v_{\text{rms}} = \sqrt{\langle v^2 \rangle} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3 \times (1.38 \times 10^{-23}) \times (293)}{28 \times (1.66 \times 10^{-27})}} = 510 \text{ m/s}$$

This speed is comparable to, but somewhat greater than, the speed of sound. This is not surprising since sound is a pressure wave in air, propagating by collisions among the air molecules. Like a message being passed from person to person, the pressure disturbance in air (the sound wave) would be expected to travel at about the speed of the messengers, the air molecules.

Example 15.2

What is the rms average speed of a diatomic bromine molecule (Br_2 , atomic number $Z = 35$) in air at room temperature?

By the equipartition theorem, the average kinetic energies of bromine molecules and N_2 molecules are identical, at the same temperature.

$$\left(\frac{1}{2}mv^2\right)_{\text{Br}} = \left(\frac{1}{2}mv^2\right)_{\text{N}} = \frac{3}{2}kT$$

We can then set up the following ratio:

$$\frac{v_{\text{Br}}}{v_{\text{N}}} = \sqrt{\frac{m_{\text{N}}}{m_{\text{Br}}}} = \sqrt{\frac{14}{70}} = 0.447$$

Using the result from Example 15.1, we have

$$v_{\text{Br}} = 0.447v_{\text{N}} = 0.447 \times (510 \text{ m/s}) = 228 \text{ m/s}$$

At any given temperature, the heavier bromine molecules travel more slowly, on average.

15.3 The Boltzmann Factor

We now come to the central result of statistical mechanics, which, oddly enough, has no standard name. We will call it the Boltzmann relation,* a name used in some textbooks. The Boltzmann relation was established before the development of quantum mechanics; however, we will give the quantum version of this equation because it is actually simpler than the original classical version. We begin by stating this relation without proof; in Section 15.5, we will explore its derivation.

For a system in equilibrium at temperature T , the probability that the system is in a particular quantum state i , a *particular microstate*, with energy E_i , is proportional to $e^{-E_i/kT}$:

$$P(\text{state } i) = C e^{-E_i/kT} \quad (15.2)$$

where P represents probability, and C is a constant of proportionality.

The factor $e^{-E_i/kT}$ is called the **Boltzmann factor**, after Ludwig Boltzmann, one of the great pioneers of kinetic theory and statistical mechanics. The constant C is determined by the following normalization argument: The system must certainly occupy *one* of the possible states i . Thus the sum over all i of the probabilities (15.2) must equal 1.

$$\sum_i P(\text{state } i) = C \sum_i e^{-E_i/kT} = 1$$

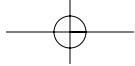
and hence,

$$C = \frac{1}{\sum_i e^{-E_i/kT}}$$

We can then rewrite (15.2) as the **Boltzmann relation**

$$P(\text{state } i) = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}} \quad (15.3)$$

*We cannot call it the “Boltzmann equation” because that phrase is already used to refer to an equation in transport theory. Boltzmann was prolific.



500 Chapter 15 • Statistical Mechanics

Notice the use of two different indexes (i and j) in this equation to avoid confusion: We consider a *particular* state i , but we sum over *all* states j . The sum

$$\sum_j e^{-E_j/kT} \quad (15.4)$$

depends on the energies E_j and is called the **partition function** because it describes how the total energy of the system is *partitioned* among the various available states. Notice that the partition function is a dimensionless number. This number turns out to be proportional to the number of different quantum states that the system is likely to be found in. See Problem 15.12.

The great power of the Boltzmann relation lies in its generality. The “system” under study might be a hydrogen atom or a neutron star, a block of metal or one electron in that metal, a mole of gas molecules in a container or one molecule in that gas. For this relation to apply, only two conditions must be met: The system must be in thermal equilibrium at temperature T , and the system must contain a definite number of particles. For systems in which the number of particles can change, a modified version of (15.3) must be used, a version beyond the scope of our discussion.*

Since the Boltzmann factor $e^{-E/kT}$ is a monotonically decreasing function of E , a particular state of higher energy is always less likely to occur than a given state of lower energy, at a given temperature. However, it does not follow that higher energies are less likely than lower energies. The probability $P(\text{state } i)$ that a system is in a quantum state i with energy E_i is not the same as the probability $P(\text{energy } E_i)$ that the system has energy E_i . As we have seen before, quantum states can be *degenerate*; that is, there can be several states with the same energy. In this case, the probability $P(E)$ that the system has energy E is the sum of the probabilities that the system is in any of the states with that energy.

$$P(E) = \sum_{\substack{\text{states } i \\ \text{with energy } E}} P(i) \quad (15.5)$$

But all the states i with the same energy E have the same probability $P(i)$, given by (15.3), so the Boltzmann relation can be rewritten

$$P(E) = g(E)P(i) = \frac{g(E)e^{-E/kT}}{\sum_j e^{-E_j/kT}} \quad (15.6)$$

where $g(E)$ is the degeneracy of the energy level E .

An analogy with playing cards may be useful. Consider the 52 cards in a deck of cards to represent 52 quantum states, and the 4 suits (hearts, clubs, spades, and diamonds) to be 4 different energies. The degeneracy of each energy is 13 (ace, two, ··· jack, queen, king). The probability of drawing a particular card, say, five of hearts, is $1/52$, but the probability of drawing a heart is

$$P(\text{heart}) = 13 \times \frac{1}{52} = \frac{1}{4}$$

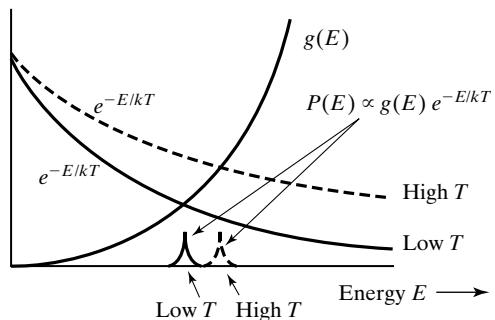
an example of $P(E) = g(E)P(i)$.

*In cases where the number of particles is not fixed, but is a variable number N , one must introduce a *chemical potential* μ , and the Boltzmann factor becomes $e^{(N\mu - E)/kT}$ — but this is a topic for a later course.

Ludwig Edward Boltzmann
(1844–1906, Austrian)



Boltzmann developed the kinetic theory of gases and invented the science of statistical mechanics. At a time when many scientists did not believe in atoms, Boltzmann's work was controversial and it was subjected to strong attacks, as well as great praise, throughout his career. He was known as an inspiring lecturer, and sometimes his classes had to be moved to larger and larger auditoriums to accommodate growing crowds of listeners. Suffering from episodes of depression all his life, he finally committed suicide, in part because he could not bear the criticism from his colleagues. Engraved on his tombstone is the equation $S = k \ln W$ (Boltzmann used the symbol W for the multiplicity g).

**FIGURE 15.2**

A highly schematic view of the competition between two factors in the Boltzmann relation (15.6).

In physical systems the degeneracy factor $g(E)$ that appears in (15.6) is almost always an increasing function of energy. For instance, in Section 15.8, we will show that for the case of a single particle in a 3-D quantum gas, $g(E) \propto E^{1/2}$. In Section 15.5 we will see that for many-particle systems, the degeneracy $g(E)$ grows very rapidly, typically $g(E) \propto E^N$, where N is the number of particles. While $g(E)$ is an increasing function of energy, the Boltzmann factor $e^{-E/kT}$ is a rapidly decreasing function of energy. The most probable energy of a system at temperature T is determined by the competition between these two factors, as indicated schematically in Fig. 15.2. When the temperature is increased, the Boltzmann factor curve in Fig. 15.2 moves up to the dashed position, resulting in a higher most probable energy for the system. Higher temperatures mean higher energies.

Example 15.3

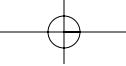
Consider a gas of hydrogen in equilibrium at room temperature, $T \approx 293$ K. What is the ratio of the number of atoms in the first excited state to the number of atoms in the ground state?

The number ratio is the ratio of the probabilities given by (15.6). In this case the “system” is a single hydrogen atom. The ground-state energy is $E_1 = -13.6$ eV; the first excited-state energy is $E_2 = (E_1/4) = -3.4$ eV. At room temperature, $kT = 0.0252$ eV. The degeneracy of the ground state is $g(E_1) = 2$ since the two spin states of the electron have equal energy; the degeneracy of the first excited state is $g(E_2) = 8$ since there is a single $2s$ state, three $2p$ states, and a factor of 2 from spin. When we take the ratio of the probabilities, the partition function factor, $\sum_j e^{-E_j/kT}$, cancels out, leaving

$$\frac{P(E_2)}{P(E_1)} = \frac{g(E_2)e^{-E_2/kT}}{g(E_1)e^{-E_1/kT}} = \frac{g(E_2)}{g(E_1)}e^{-(E_2-E_1)/kT} = \frac{8}{2}e^{-405} \approx 10^{-176}$$

This is such a fantastically small number that if the observable universe were filled with hydrogen gas at 1 atm of pressure, then at room temperature there would not be even one hydrogen atom in the first excited state.

In problems such as this one, where the Boltzmann factor is used to compute relative probabilities, it is expedient to set up a ratio so that the partition function cancels out. In all but the simplest cases, the partition function is quite difficult to compute exactly, so we avoid that task wherever possible.



502 Chapter 15 • Statistical Mechanics

Example 15.4

In hydrogen gas, at what temperature is the ratio of the number of atoms in the first excited state to the number of atoms in the ground state equal to 1/100; that is, how hot does hydrogen gas have to be for a significant fraction of the atoms to be in excited states?

Proceeding as in Example 15.3,

$$\frac{P(E_2)}{P(E_1)} = \frac{g(E_2)}{g(E_1)} \frac{e^{-E_2/kT}}{e^{-E_1/kT}} = 4e^{-\Delta E/kT} = \frac{1}{100}$$

therefore, $\Delta E/kT = \ln(400)$, or

$$kT = \frac{\Delta E}{\ln(400)} = \frac{10.2 \text{ eV}}{5.99} = 1.70 \text{ eV}$$

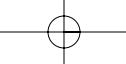
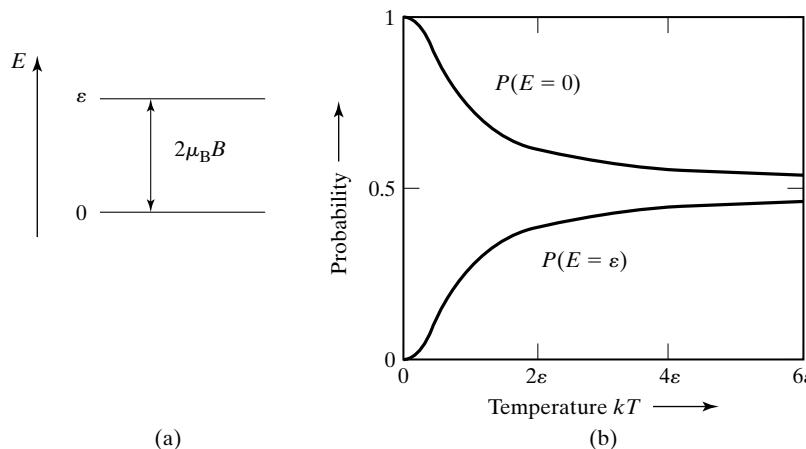
This corresponds to a temperature of $T \approx 20,000$ K. The universe was at this temperature about 100,000 years after the “big bang.” At that time, in addition to excited hydrogen atoms, there was a substantial density of free electrons and protons. These charged particles scatter light strongly, and so during that epoch the universe was opaque; light could not travel far before scattering. At about 300,000 years after the big bang, when the universe had cooled to 3000 K, virtually all free electrons and protons had combined to form hydrogen atoms in the ground state. Ground-state hydrogen atoms are completely transparent to low-energy photons (since photons of energy less than 10.2 eV cannot be absorbed), so light could travel freely without scattering. Today, some 15 billion years later, astronomers have observed that space is filled with microwave background radiation. This photon gas, which is at a temperature of 3 K, is the Doppler-shifted light from that time 300,000 years after the big bang, when matter and radiation “decoupled.”

Example 15.5

Consider a particle with spin half and a magnetic moment μ in an external magnetic field \mathbf{B} . Two orientations of the moment with respect to the field are allowed, corresponding to spin up and spin down. Find the probability that the moment is aligned with the field, in terms of the temperature T and field B . What is this probability at room temperature (300 K) in a field of $B = 1.0$ T?

As described in Section 9.6, an electron in an external magnetic field has two energy states: a ground state with the moment aligned with the field and a first excited state with the moment anti-aligned with the field. The separation of the energy levels is $\varepsilon = 2\mu_B B$ where μ_B is the Bohr magneton. This is an example of a two-level system, a system with exactly two states, a ground state and an excited state. A two-level system is one of the few systems for which it is easy to write down the partition function (15.4) exactly. If we set the zero of energy at the ground-state energy, then the two levels have energies 0 and ε [see Fig. 15.3(a)], and the partition function becomes

$$\sum_j e^{-E_j/kT} = 1 + e^{-\varepsilon/kT}$$

Section 15.4 • Counting Microstates: The Equal-Probability Hypothesis **503****FIGURE 15.3**

(a) The energy spectrum of an electron in a magnetic field, an example of a two-level system.
 (b) Probabilities that the electron is found in the ground state and in the excited state, as functions of kT .

The probabilities of the two states are then, according to (15.3)

$$P(E = 0) = \frac{1}{1 + e^{-\varepsilon/kT}} \quad \text{and} \quad P(E = \varepsilon) = \frac{e^{-\varepsilon/kT}}{1 + e^{-\varepsilon/kT}} = \frac{1}{e^{\varepsilon/kT} + 1}$$

These two probabilities are plotted in Fig. 15.3(b). The upper curve is the probability $P(E = 0)$ that the system is in the ground state, with the magnetic moment aligned with the field; the lower curve is the probability that the system is in the excited state. In the high-temperature limit ($kT \gg \varepsilon$), both probabilities approach 0.5 and the system is equally likely to be in either state.

At $B = 1.0$ T, the energy-level separation is $2\mu_B B = (2)(5.79 \times 10^{-5}$ eV/T)(1T) = 1.16×10^{-4} eV, and the ground-state probability at $T = 300$ K ($kT = 300$ K ($kT = 0.252$ eV)) is

$$P(E) = \frac{1}{1 + e^{-\varepsilon/kT}} = \frac{1}{1 + e^{-2\mu_B B/kT}} = \frac{1}{1 + e^{-(1.16 \times 10^{-4})/(0.0252)}} = 0.5012$$

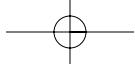
At room temperature, we are in the high-temperature regime ($kT \gg 2\mu_B B$) and the populations of the two levels are almost equal. There is a preference for the moments to be aligned with the B field, but the degree of alignment (or polarization) is only about 1 part in 500.

15.4 Counting Microstates: The Equal-Probability Hypothesis

The whole of statistical mechanics, including the Boltzmann relation, rests upon a single axiom — an axiom that, like the Boltzmann relation, has no standard name. We will call it the

EQUAL-PROBABILITY HYPOTHESIS

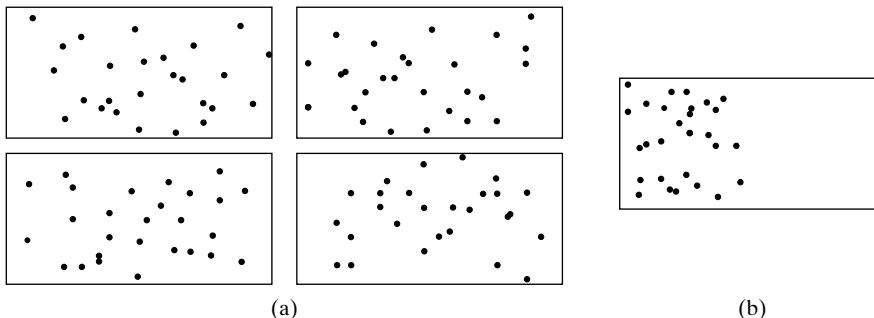
For an isolated system (that is, one with constant energy, constant number of particles, and so forth.) in thermal equilibrium, all microstates accessible to the system occur with equal probability.



504 Chapter 15 • Statistical Mechanics

FIGURE 15.4

Some microstates of a gas in a container. The seemingly peculiar state (b) is just as likely as any specific one of the states in (a), but there are far fewer states of the type (b) than of the type (a).



By *accessible* microstates, we mean those microstates that are consistent with the macroscopic constraints on the system — constraints such as the total energy, the total number of particles, and, in the case of a gas, the volume and shape of the container. These constraints define the *macrostate*. Thus, another way to state this hypothesis is this: All microstates corresponding to a given macrostate occur with equal probability. There is no known proof of this assumption; we believe it is true because it leads to many predictions that are all verified by experiment. We will see in the next section how this hypothesis leads to the Boltzmann relation.

As an example of the equal probability of accessible microstates, consider a gas of N particles in a container of volume V . The macrostate is given by (N, E, V) , where E is the total energy of the system. Classically, a microstate is specified by the positions and velocities of each of the N particles. In Fig. 15.4(a) four microstates are shown all corresponding to the same macrostate (N, E, V) . There is no reason for the system to prefer any of these states over the others, and they all occur with equal probability. Now consider the microstate in Figure 15.4(b), also a state with the same (N, E, V) as the others. This state appears unusual because all of the particles are in the left half of the container. However, strange as it may seem, this “left-side-only” microstate is just as likely as any other *particular* microstate. The reason we never observe gas molecules spontaneously clustering on one side of a container is because the microstates with the particles evenly dispersed vastly outnumber the microstates in which the particles cluster together. If a microstate is chosen at random, there is an overwhelming likelihood that it will be one with roughly uniform density.

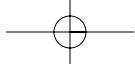
To get some feeling for microstates and macrostates, we again resort to a playing-card analogy. There are a total of $T = 52$ cards in a deck, but only $N = 5$ cards in a poker hand. Specifying N defines the macrostate; in this analogy the macrostate is “a hand of 5 cards” — somewhat akin to specifying “5 mol of gas at standard temperature and pressure.”* A microstate is a *particular* hand, for instance the hand

$$(3\spadesuit, J\heartsuit, 10\clubsuit, 4\clubsuit, 6\clubsuit)$$

Note that in specifying a hand, the order of the cards is irrelevant. The hand

$$(J\heartsuit, 3\spadesuit, 10\clubsuit, 4\clubsuit, 6\clubsuit)$$

*Of course, in poker there is no clear analogue to temperature or pressure. We cannot push this analogy too far.

Section 15.4 • Counting Microstates: The Equal-Probability Hypothesis **505**

is the same microstate as the first hand; same cards, just different order. Likewise, in a gas of indistinguishable particles, interchanging two particles leads to the same microstate. (Indistinguishability was discussed in Ch. 10 and is further reviewed below.) There are many possible poker hands, many possible microstates, all of which are equally likely to be dealt from a well-shuffled deck. A royal flush

$$(10\clubsuit, J\clubsuit, Q\clubsuit, K\clubsuit, A\clubsuit)$$

is just as likely to occur as any *particular* “junk” hand

$$(3\spadesuit, J\heartsuit, 10\spadesuit, 4\clubsuit, 6\clubsuit)$$

The reason junk hands are found so much more often than royal flushes is that there are only 4 possible royal flush hands (corresponding to the 4 suits), but there are more than a million possible junk hands.

To compute the probability of being dealt a royal flush, we need to know how many possible microstates (hands) there are. The number or “multiplicity” g of five-card hands ($N = 5$) that can be dealt from a deck with $T = 52$ cards is given by

$$g = \frac{T!}{(T-N)!N!} \text{ (number of ways of choosing } N \text{ objects from among } T\text{)} \quad (15.7)$$

This formula, which is called the binomial coefficient, can be understood as follows: The first card dealt in the 5-card hand can be any one of 52 cards; the second card can be any one of the remaining 51, and likewise for the third, and so on. The number of different *ordered* 5-card hands is then

$$52 \times 51 \times 50 \times 49 \times 48 = \frac{52!}{47!} = \frac{52!}{(52 - 5)!}$$

But we have overcounted the number of hands, because the order of the cards is irrelevant. There are $5 \times 4 \times 3 \times 2 \times 1 = 5!$ ways of ordering 5 cards, so we must divide the expression above by $5!$ to get the number of distinct hands. The multiplicity g is then

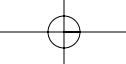
$$g = \frac{T!}{(T - N)! N!} = \frac{52!}{(52 - 5)! 5!} = 2,598,960$$

Out of all these hands, only 4 are royal flushes, so the probability of dealing a royal flush is

$$P = \frac{4}{2598960} \approx 1.54 \times 10^{-6}$$

This is a very small probability, but not as small as, say, the probability of winning a large state lottery ($P \approx 10^{-8}$) and not nearly as small as the probabilities often encountered in statistical mechanics.

There is a small probability that the air molecules around you will spontaneously cluster in a corner of the room, leaving you to suffocate as you read these words. You need not worry; the probability of this occurring, although



506 Chapter 15 • Statistical Mechanics

nonzero, is *fantastically* small, as we now demonstrate. Consider a mole of gas molecules in a container. Any molecule can be anywhere in the container; of all the possible states (position and velocity) that a single molecule can occupy, there are as many on the right half of the container as on the left half. If we choose a state, *at random*, for a molecule to occupy, the probability that the one molecule will be on the left is $P(1) = 1/2$. If we randomly choose 2 states for 2 molecules, the probability that both particles will be on the left is the same as getting 2 heads in 2 coin flips, $P(2) = (1/2)(1/2) = 1/4$. The probability of choosing N states randomly and finding all N states on the left is

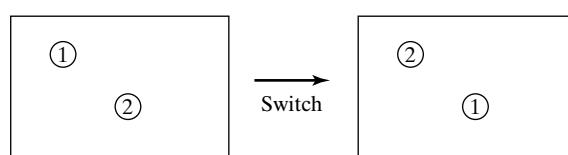
$$P(N) = \frac{1}{2^N} \quad (15.8)$$

For a mole of gas molecules, $N = 6 \times 10^{23}$, and $P(N) = 0.5^N = 10^{-N \ln 2 / \ln 10} \approx 10^{-2 \times 10^{23}}$ (see Problem 15.22). It is difficult to describe how tiny this number is. If you wanted to write this number in decimal form as 0.000000...001, you would have to write 2×10^{23} zeroes. If you could write a 100,000 zeroes per second, it would take you the age of the universe ($\approx 10^{10}$ years) to write all those zeroes.

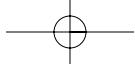
Exactly how one counts the microstates of a gas of molecules depends on whether the molecules are *distinguishable* or *indistinguishable*. Recall from Section 10.5 that in classical mechanics, identical particles are distinguishable, meaning that it is possible, in principle, to keep track of which particle is which by following their trajectories. In quantum mechanics, however, it is not possible, even in principle, to keep track of which particle is which because the wave functions of particles can overlap. We saw in Section 10.5 that this indistinguishability of identical particles has a profound effect on the symmetry of many-particle wave functions. When two indistinguishable particles in a gas have their positions interchanged, all observable properties must be unchanged; in particular, we must have $|\psi(1, 2)|^2 = |\psi(2, 1)|^2$.* All particles can be classified as either *fermions* or *bosons*, depending on whether or not the wave function changes sign under particle exchange. For fermions, $\psi(1, 2) = -\psi(2, 1)$, which leads to the Pauli exclusion principle: No two fermions can exist in the same quantum state. For bosons, $\psi(1, 2) = +\psi(2, 1)$ and there is no constraint on the number of particles that can exist in the same quantum state.

In the real world, particles obey quantum mechanics and identical particles are always indistinguishable. Interchanging the positions of indistinguishable particles leads to the *same* microstate. In classical mechanics, however, switching identical but distinguishable particles leads to a *different* (though physically identical) microstate. See Fig. 15.5. We now show that, although distinguishability has a big effect on the number of microstates, it often has no effect on computed probabilities, such as (15.8).

Once again, we compute the probability that all the gas molecules in a box will, quite by chance, cluster in the left half of the box. We repeat this



*As in Ch. 10, we use 1 and 2 to represent the appropriate coordinates for particles 1 and 2. For example, for a spin-half particle, 1 = (\mathbf{r}_1, m_{s1}) .

Section 15.4 • Counting Microstates: The Equal-Probability Hypothesis **507**

exercise to show how probabilities are computed by counting microstates and to show that in the classical regime we get the same probability regardless of whether or not the particles are distinguishable. We consider N identical particles in a sealed container and assume that the particles can move freely between the left and right halves of the container. To keep the math uncomplicated, we will model the quantum states of this system very simply: We assume that there are H different single-particle quantum states on the left half of the box, that is, states with wave functions localized on the left, and there are another H states on the right. Hence, there are a total of $2H$ different quantum states that any of the N particles can occupy (see Fig. 15.6). We also assume that $H \gg N$, which means there are many more quantum states than particles, so the particles are not crowded into a few states. (This is the *nondegenerate, classical regime* described in Section 13.11.) If the particles are fermions, then no more than one particle can occupy any of the $2H$ states. If the particles are bosons, any number of particles can exist in any of the quantum states. Here we solve the problem for the case of fermions; the boson case is left as an exercise (Problem 15.26).

The probability that all the particles will cluster on the left is the ratio of the number of ways of arranging the N particles among the H states on the left to the total number of ways of arranging the particles among all $2H$ states.

$$P(\text{left side only}) = \frac{\text{number of ways of arranging } N \text{ particles among } H \text{ states of left half}}{\text{number of ways of arranging } N \text{ particles among all } 2H \text{ states.}} \quad (15.9)$$

This equation is true because all arrangements, all microstates, occur with equal probability.

Whether the particles are distinguishable or indistinguishable affects the counting of the microstates. Let us first assume, as a nineteenth-century classical physicist would, that the particles are distinguishable. As we know, this assumption is *incorrect*, but let us proceed anyway.

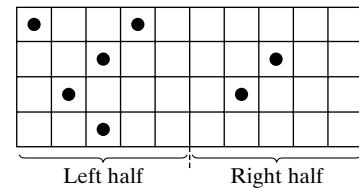
The number of ways of arranging N *distinguishable* particles among H states, with no more than one particle per state is

$$g(N, H) = \frac{H!}{(H - N)!} \quad (\text{distinguishable particles}) \quad (15.10)$$

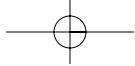
To see this, imagine placing N balls in H slots. The first ball can go into any of H slots. The second ball can go into any of the remaining $(H - 1)$ slots, and so on. The total number of arrangements is $H \cdot (H - 1) \cdot (H - 2) \cdots (H - N + 1)$, which is equal to (15.10). Similarly, the total number of arrangements of N distinguishable particles among $2H$ states is $g(N, 2H) = (2H)!/(2H - N)!$. We can now compute the ratio (15.9), and we find

$$P(\text{left side only}) = \frac{H!}{(2H)!} \frac{(2H - N)!}{(H - N)!} \quad (15.11)$$

This is the probability that all N distinguishable particles will spontaneously cluster on the left. We now show that this is also the probability that N *indistinguishable* particles will cluster on the left.

**FIGURE 15.6**

Seven particles ($N = 7$) in a box with $2H = 40$ quantum states. There are $H = 20$ states in the left half of the box and another H states in the right. The counting of microstates in this system depends on whether the particles are distinguishable or not.



508 Chapter 15 • Statistical Mechanics

We repeat the calculation with *indistinguishable* particles. The number of ways of arranging N indistinguishable particles among H states, with no more than one particle per state, is

$$g(N, H) = \frac{H!}{N!(H - N)!} \quad (\text{indistinguishable particles}) \quad (15.12)$$

This is the same as (15.7), the poker hand multiplicity equation — N cards from a 52-card deck, or N particles among H states, the counting argument is the same. Note that the only difference between equations (15.10) and (15.12) is an extra factor of $1/N!$ in (15.12) because there are $N!$ ways of ordering N particles. For the case of indistinguishable particles, these $N!$ arrangements all correspond to the same state. In exactly the same way, there is an extra factor of $1/N!$ in the probability $P(N, 2H)$.

When we now recompute the ratio (15.9), we have an extra factor of $1/N!$ in both numerator and denominator. The extra factors cancel and we get exactly the same answer as before, Eq. (15.11). This formula for the probability looks rather formidable, but it can be shown that it reduces to the simpler formula (15.8) in the limit $H \gg N \gg 1$. (See Problem 15.25.) We conclude that in the low-density classical regime, in which $H \gg N$, the probability is not affected by the distinguishability of the particles. (There are situations, however, in which distinguishability does make a difference. One such case is treated in Problem 15.36.)

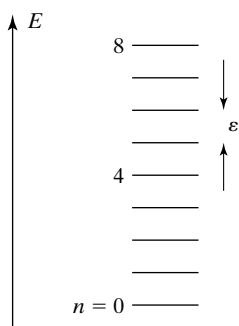


FIGURE 15.7

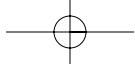
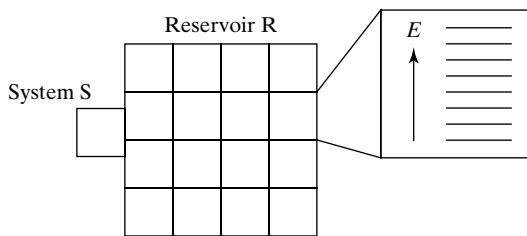
The energy spectrum of a simple quantum system.

15.5 The Origin of the Boltzmann Relation*

*This section contains rather difficult arguments and can be omitted if you are satisfied to accept the Boltzmann relation (15.3) without proof.

Our goal in this section is to understand how the Boltzmann relation arises from the equal-probability hypothesis. Although the Boltzmann relation applies to any quantum system, we will consider the special case of a system with a very simple energy spectrum: a set of uniformly-spaced, nondegenerate energy levels given by $E_n = n\epsilon$, where ϵ is a fixed energy and n is any non-negative integer: $n = 0, 1, 2, \dots$. (See Fig. 15.7.) This happens to be the energy-level spectrum of the 1-D simple harmonic oscillator described in Section 7.9, but the details of the system are irrelevant here; all we need to know is the energy spectrum.

We assume that this model microscopic system is in thermal equilibrium with a large thermal reservoir at temperature T . By **thermal reservoir**, we mean some large macroscopic system with a very large number of degrees of freedom, with which our microscopic system can exchange energy. Temperature is a macroscopic concept; it makes no sense to talk about the temperature of an *isolated* microscopic system, such as a single atom. The temperature of a microscopic system is the temperature of the thermal reservoir with which it is in equilibrium. The thermal reservoir is assumed to be so large that the exchange of energy with our micro-system has negligible effect on the reservoir's temperature, that is, on its average energy per degree of freedom. Finally, we assume that the reservoir and the system are thermally isolated from everything else. Hence, the sum of the energies of the reservoir and our system is fixed. It is this model “universe” of (system + reservoir) to which we will apply the equal-probability hypothesis.

Section 15.5 • The Origin of the Boltzmann Relation **509****FIGURE 15.8**

A system S in contact with a thermal reservoir R . The reservoir is a collection of identical quantum systems, each with the same simple energy spectrum.

The thermal reservoir, also called a “heat bath,” can be modeled as a collection of a large number N of microscopic quantum systems in thermal contact. To keep things as simple as possible, we assume that each of the N quantum systems making up the reservoir is identical to our original microscopic system, with the simple energy spectrum $E_n = n\epsilon$. Our universe of (system + reservoir) then consists of $N + 1$ quantum systems, all in thermal contact, freely exchanging energy among themselves, as in Fig. 15.8. Let us denote the total energy of the (system + reservoir) by $m\epsilon$, where m is some integer, equal to the number of energy quanta of size ϵ . Since the total energy of the (system + reservoir) is fixed, when the system has energy $E_S = s\epsilon$, the reservoir must have energy $E_R = r\epsilon = (m - s)\epsilon$.

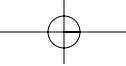
When the system is in some *particular* microstate s with energy E_S , the reservoir can be in *any* microstate with energy $E_R = (m - s)\epsilon$. According to the equal-probability hypothesis, all microstates of the (system + reservoir) are equally likely; hence, the probability that the system is in state s is proportional to the number of microstates of the (system + reservoir) that have the system in the state s . This is just the number of microstates of the reservoir that have energy $E_R = (m - s)\epsilon$. We conclude that the probability that the system is in state s is proportional to the degeneracy (the multiplicity) of the reservoir when it has energy $E_R = (m - s)\epsilon$. Or, in symbols, $P[\text{system is in particular state } s] = P[\text{reservoir is in any state with energy } (m - s)\epsilon]$,

$$P(s) \propto g_R(m - s) \quad (15.13)$$

The key point here is to realize that the equal-probability hypothesis applies to the (system + reservoir); we cannot apply it to the microscopic system alone since the microsystem is not an isolated system; it exchanges energy with the heat bath. Our job now is to show that Eq. (15.13) is equivalent to the Boltzmann relation (15.2).

The multiplicity of the reservoir $g_R(r)$ grows very rapidly with the number r of energy quanta. For example, if there are $N = 100$ systems in the reservoir, and only one ($r = 1$) quantum of energy, then the number of reservoir microstates is $g(r = 1) = 100$ corresponding to the single energy quantum being in any one of the 100 systems. If there are $r = 2$ energy quanta, then there are $(100 \cdot 99)/2$ different microstates with two of the 100 systems containing one quantum each, plus another 100 microstates with both quanta in one of the 100 systems. Hence, $g(2) = (100 \cdot 99)/2 + 100 = 5050$. The general formula for the multiplicity, when there are r indistinguishable quanta distributed among N distinct systems, is

$$g(r) = \frac{(N + r - 1)!}{r!(N - 1)!} \quad (15.14)$$



510 Chapter 15 • Statistical Mechanics

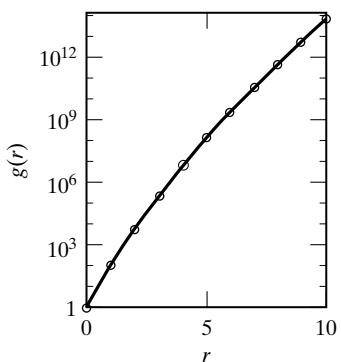


FIGURE 15.9

The number of ways of placing r identical objects in $N = 100$ containers.

The proof of this formula, which is very similar to the proof of (15.7), is explored in Problem 15.29. A graph of $g(r)$ for the case $N = 100$ is shown in Figure 15.9. Note that $g(r)$ grows *extremely* rapidly with increasing r . In the limit $r \gg N$, which corresponds to the high-temperature limit where there are many energy quanta per system, Eq. (15.14) becomes approximately $g(r) \approx r^N \propto E^N$, where $E = r\epsilon$ is the total energy of the system. (See Problem 15.30.)

We can now understand an important qualitative feature of the Boltzmann factor, $e^{-E/kT}$. According to the Boltzmann relation (15.2), a particular state i of lower energy E_i is always more likely to occur than a particular state j of higher energy E_j . When the microscopic system has low energy (small s), the reservoir has more energy [larger $(m - s)$]. When the reservoir has more energy, there are more ways to divide up that energy, resulting in a larger number of accessible microstates, a greater multiplicity, and a greater probability. Therefore, low-energy states of the system are more probable.

With the multiplicity of the reservoir (15.14) known, the probability (15.13) that the system is in a given state s can be written as

$$P(s) \propto g_R(m-s) = \frac{[N + (m-s) - 1]!}{(m-s)!(N-1)!} \quad (15.15)$$

We leave it as an exercise (Problem 15.31) to show that, in the limit of large $m \gg s$ (a big reservoir) and $m \gg N$ (a high temperature), (15.15) can be rewritten as

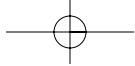
$$P(s) \propto e^{-s/(m/N)} \quad (15.16)$$

This is none other than the Boltzmann relation (15.2) with the system energy $E = s\epsilon$ and temperature $kT = (m/N)\epsilon$. To see how temperature enters, recall that according to the equipartition theorem, the energy per degree of freedom is $\frac{1}{2}kT$. For the N oscillators of the reservoir, there are $2N$ degrees of freedom (two degrees of freedom for each simple harmonic oscillator). The total energy available is $m\epsilon$, so we have

$$\frac{1}{2}kT = \frac{\text{energy}}{\text{number of degrees of freedom}} = \frac{m\epsilon}{2N}$$

or $kT = (m/N)\epsilon$. We have imposed the condition $m \gg N$, so that the mean energy per oscillator in the thermal reservoir is large compared to the energy quantum ϵ , in which case the energy can be regarded as a quasi-continuum and the classical equipartition theorem applies.*

*The Boltzmann relation is true always, even at very low temperatures where the equipartition theorem does not apply. However, when the equipartition theorem does not apply, we need a more general definition of temperature than the one we have given in Eq. (15.1). The more general definition relates temperature to the entropy S and the internal energy U , namely $1/T = \partial S/\partial U$. [See, for instance, *Thermal Physics* (Ed.) by C. Kittel and H. Kroemer, W H Freeman & Co., 1980.]



15.6 Entropy and the Second Law of Thermodynamics*

*This section assumes some previous exposure to thermodynamics and can be omitted.

The multiplicity, or degeneracy, g of a system is directly related to its *entropy* S . In thermodynamics the change in the entropy of a system is defined by the relation

$$dS = \frac{dQ}{T} \quad (15.17)$$

where dQ is the heat added or removed from the system and T is the absolute temperature in kelvins. This equation applies only to *reversible* processes, that is, processes in which the whole system remains in thermal equilibrium so that it has a well-defined temperature T . Before the development of statistical mechanics, the thermodynamic entropy defined by (15.17) was a mysterious quantity. One could compute its value (or at least *changes* in its value), but one could not say what it was microscopically. This uncomfortable situation was reminiscent of the incomplete understanding of temperature that existed before the advent of kinetic theory and the atomic hypothesis.

It was Boltzmann who first saw clearly the connection between macroscopic thermodynamics and microscopic statistical mechanics. Boltzmann showed that the thermodynamic entropy S of a macrostate is related to the natural logarithm of the multiplicity g by the equation

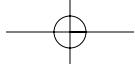
$$S = k \ln(g), \quad (15.18)$$

where k is Boltzmann's constant, so named in honor of the intellectual breakthrough represented by this equation. The multiplicity g is the number of microstates available to the system. With this equation, we can now understand better why entropy is often called the "disorder" of a system. A disordered state, like a junk hand, has a larger multiplicity g and hence a larger entropy than a highly ordered state, like a royal flush.

In light of equation (15.18), we can also better understand one of the most profound achievements of nineteenth-century physics, the second law of thermodynamics. The second law, also called the law of increase of entropy, states that the entropy of a closed system never decreases; it can only increase or stay constant. For instance, when an internal constraint, such as a barrier between the two halves of a container of gas, is removed, the entropy always increases. From (15.18) we see that this is because the number of accessible microstates always increases when constraints are removed.

Example 15.6

A well-insulated container is divided in half by a thin barrier. In the left half of the container is an ideal gas of N particles at some temperature T ; the right half of the container is empty. The barrier breaks, doubling the volume of the gas from V_0 to $2V_0$, as shown in Figure 15.10. What is the increase in entropy of the system resulting from this adiabatic expansion of the gas? (An **adiabatic** process is one in which no heat is added to or removed from the system.)



512 Chapter 15 • Statistical Mechanics

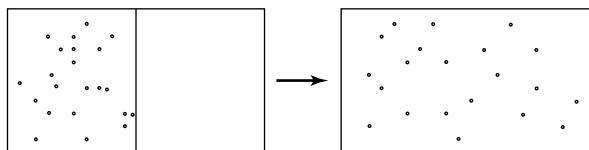


FIGURE 15.10

Adiabatic expansion of a gas.

This problem can be solved in two ways: by counting microstates and using (15.18), or by a thermodynamic calculation using (15.17).

Method I, counting microstates. If g_0 is the original multiplicity of the system before the barrier broke, then after the volume of the system doubles, the new multiplicity is $2^N g_0$. We show this by the following classical argument: If the gas consisted of only a single particle, the number of microstates g would double because for every single-particle state on the left, there is a corresponding state on the right. If the gas consisted of two particles, the multiplicity would increase by a factor $2^2 = 4$, since each particle can be in either a right state or the corresponding left state and there are 2×2 possible combinations. With N particles, the multiplicity increases by 2^N , from g_0 to $2^N g_0$. The final entropy is then

$$S = k \ln(2^N g_0) = k \ln(g_0) + k \ln(2^N) = k \ln(g_0) + Nk \ln(2)$$

Thus, the increase in entropy is

$$\Delta S = Nk \ln(2) \quad (15.19)$$

Method II, thermodynamics. In the free expansion of the gas, no heat was added or removed ($dQ = 0$) and no work was done ($dW = 0$), so the internal energy U of the gas remains unchanged ($dU = dQ + dW = 0$). For an ideal gas, $U = (3/2)NkT$, and so the temperature is also unchanged. The free expansion of a gas is a nonreversible process, but equation (15.17) $dS = dQ/T$ applies only to reversible (equilibrium) processes. So to use (15.17), we must imagine some reversible process that results in the same final state of the system. One such process is this: The barrier is slowly moved right, like a piston head, and at the same time the temperature of the system is held fixed by slowly adding heat to the system to compensate for the work done on the piston ($dQ = -dW$, so $dU = 0$). During this reversible, constant-temperature expansion, $dQ = -dW = p dV$, and so

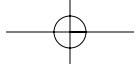
$$\frac{dQ}{T} = \frac{p}{T} dV$$

whence,

$$\Delta S = \int \frac{dQ}{T} = \int_{V_0}^{2V_0} \frac{p}{T} dV = \int_{V_0}^{2V_0} \frac{Nk}{V} dV = Nk \ln\left(\frac{2V_0}{V_0}\right) = Nk \ln(2)$$

in agreement with the result of Method I, (15.19). (The third equality follows from $pV = NkT$.)

The second law of thermodynamics gives direction to the arrow of time. In this world, the sequence of events is always more order, then less order —

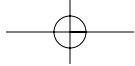
Section 15.7 • The Quantum Ideal Gas — a Many-Particle System **513**

lower entropy, then higher entropy. Buildings eventually fall down and become rubble; we die and become dust. But one never sees the reverse; a field of dirt never spontaneously rearranges its molecules to form a house or a human. Such a miracle could occur without violating conservation of energy or momentum, just as a large pocket of vacuum could spontaneously occur in a gas, or a ball could, of its own accord, roll up hill by becoming cooler than its surroundings. But these incredible events *would* violate the second law. They would result in a decrease in the entropy of the universe and are therefore fantastically unlikely to occur — not impossible, but so unlikely that they would never occur in a trillion trillion trillion (put in as many trillions as you like) years. Of all the microstates of the field of dirt, there are many many more “rubble” states than there are “house” states, many more junk hands than royal flushes. If you pick a state at random, you are overwhelmingly likely to find rubble.

So the second law of thermodynamics says that decay is the way of the world; all things, left to themselves, eventually become more disordered. Heterogeneous structures become more uniform, sources of heat cool, houses and humans turn to dust. Why is it then, that here on earth, we are surrounded by complex, highly ordered structures, structures of low entropy, like computers and cities and books and people? It is because these structures did not arise from spontaneous fluctuations in isolated systems. They were built at great expense. In order to reduce the entropy in some part of the universe, the entropy elsewhere must be increased, for the total entropy of the universe can only increase. In our corner of the universe, the Sun is the local entropy factory, becoming more disordered at a terrific rate. Rest-mass energy, tightly locked in compact hydrogen nuclei, is being converted into gamma rays by thermonuclear reactions and then into visible photons by scattering from ions. These photons pour out into space, spreading the Sun’s energy over an immense volume. There are many many more ways to spread the immense number of photons about in space than there are ways to arrange the few hydrogen nuclei that produced those photons in the Sun’s interior. And so the multiplicity and entropy of the Sun’s energy are increasing all the time. However, as the visible light photons produced on the 6000 K surface of the Sun spread out through space, some are captured by photosynthesizing plants, which feed people, who build houses. Most of the energy taken in by plants and people is eventually expelled as heat in the form of low-energy infrared photons corresponding to a temperature near 300 K. Thus, a few high-energy visible photons moving away from the Sun are turned into many, low-energy infrared photons moving in random directions, resulting in an overall entropy increase. Biological structures and all the other wonders of the world are thermodynamically possible because they are *open* systems, *metabolizing* systems, taking in low entropy structures and expelling high entropy ones. You can create order in some places, but only by making more disorder elsewhere. We exist because we leave chaos in our wake.

15.7 The Quantum Ideal Gas — a Many-Particle System

As an example of a macroscopic system that can, in principle, be described in complete microscopic detail, we consider the quantum-mechanical problem of an ideal gas: N *noninteracting* particles in a cubical box. The problems of a single particle in a 2-D square box and in a 3-D cubical box were treated in Section 8.3. We begin by reviewing briefly the results of that section.



514 Chapter 15 • Statistical Mechanics

The Schrödinger equation for a single particle of mass M in a three-dimensional box is

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \frac{2M}{\hbar^2}[U - E]\psi \quad (15.20)$$

where the potential $U(x, y, z)$ is zero inside the cubical region $0 \leq x \leq a$, $0 \leq y \leq a$, $0 \leq z \leq a$, and is infinite everywhere outside. The wave function $\psi = \psi(x, y, z)$ must be zero at the boundaries of the box, since ψ must be continuous and it is zero outside the box. This problem is solved by the technique of separation of variables, in which one seeks solutions of the form

$$\psi(x, y, z) = X(x)Y(y)Z(z) \quad (15.21)$$

Plugging the assumed form (15.21) into equation (15.20) and performing some algebra (Problem 15.37), we find the solution

$$\psi(x, y, z) = A \sin\left(\frac{n_x \pi x}{a}\right) \sin\left(\frac{n_y \pi y}{a}\right) \sin\left(\frac{n_z \pi z}{a}\right) \quad (15.22)$$

where the quantum numbers n_x , n_y , and n_z are any positive integers ($n_x = 1, 2, 3, \dots$ etc.) and the constant A is determined by the normalization condition. Corresponding to these solutions are the allowed energies,

$$E = E_{n_x, n_y, n_z} = \frac{\hbar^2 \pi^2}{2 M a^2} (n_x^2 + n_y^2 + n_z^2) \quad (15.23)$$

Some economy of notation is achieved by defining a “ k value” associated with each value of n ,

$$k_x = \frac{n_x \pi}{a} \quad (15.24)$$

and similarly for k_y and k_z . Our wave functions and their energies are then written,

$$\psi(x, y, z) = A \sin(k_x x) \sin(k_y y) \sin(k_z z),$$

with energy

$$E = \frac{\hbar^2}{2 M} (k_x^2 + k_y^2 + k_z^2) \quad (15.25)$$

Each of our solutions to the single-particle Schrödinger equation is completely described by the set of three quantum numbers (n_x, n_y, n_z) or equivalently, by the three k values (k_x, k_y, k_z) . Each allowed set of k values (k_x, k_y, k_z) can be represented by a point in an abstract three-dimensional space called k space, which we represent with three mutually perpendicular axes k_x , k_y , and k_z , as shown in Figure 15.11. In k space the allowed points (k_x, k_y, k_z) form a cubic array of evenly spaced points. According to (15.24), as we move along a direction parallel to the k_x axis, the allowed points have the k_x values $\pi/a, 2\pi/a, 3\pi/a$, and so forth so the spacing of adjacent points is

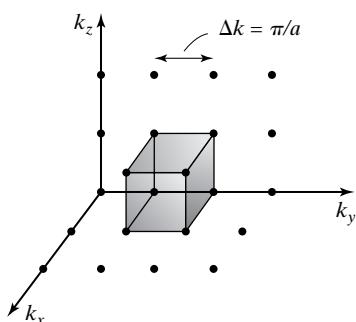
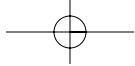


FIGURE 15.11

The states of a particle in a cubical box of side a are identified by points in k space. The allowed points form a cubic array with nearest-neighbor distance equal to π/a .

Section 15.7 • The Quantum Ideal Gas — a Many-Particle System **515**

$\Delta k = \pi/a$. Notice that in k space, “distance” has the units of $1/(\text{length})$. One can think of k_x , k_y , and k_z as the three components of a vector \mathbf{k} , whose magnitude $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ is the “distance” from the point (k_x, k_y, k_z) to the origin. Notice also that, according to (15.25), the energy of an allowed state is proportional to this distance squared, $k_x^2 + k_y^2 + k_z^2$, so that points further from the origin correspond to states of higher energy.

It may seem strange to refer to these k 's as if they were real. It may even seem strange that we use the term k , rather than some name. As mentioned in Section 6.6, physicists sometimes refer to k as “the wave number,” but most often, it is simply called k . Physically, the k value is related to the momentum of a particle by $p = \hbar k$ and it is related to the de Broglie wavelength of the particle by $k = 2\pi/\lambda$. Large k means short wavelength, large momentum, and large energy. A particular point in the k space array represents a particular solution to the single-particle Schrödinger equation. Like all things new, it takes some getting used to.

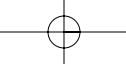
So far, we have considered only a single particle in our cubical box. What now happens when we place N particles in the box and form a many-particle system? To answer this question, we must begin by assuming that the interactions between the particles are sufficiently weak that the particles can be regarded as noninteracting. This means that the wave function of a single, given particle in the system of many particles can still be written as (15.22) with an energy given by (15.23). We assume that the presence of other particles in other states has negligible effect on this single-particle description. Perhaps surprisingly, this *independent-particle approximation* is accurate in many real, physical systems. For instance, in Chapter 13, we saw that conduction electrons in a metal behave like a gas of noninteracting particles. This is quite surprising since electrons interact via the Coulomb force, which is both strong and long range. However, in a metal, the negatively charged conduction electron gas coexists with an array of positively charged ions, and the two signs of charges effectively cancel.

If we have many noninteracting particles in our box, each particle in the system is described by a point in the lattice of allowed k -space points. A microstate of this many-particle system is a set of occupied k -space points, or equivalently, a complete set of quantum numbers $(n_{1x}, n_{1y}, n_{1z}, n_{2x}, n_{2y}, n_{2z}, \dots)$ describing the state of every particle. Which states are likely to be occupied depends on the temperature and, possibly, on the particle type — fermion or boson. We consider first the high-temperature case.

If the temperature of our system of noninteracting particles is sufficiently high, and the density of particles is sufficiently low, then our gas is in the nondegenerate, classical regime, described in Section 13.11 (see Problem 15.40). In this regime the occupied k -space points are widely scattered among a nearly empty array. As shown in Figure 15.12(a), in k space, the particles form a sparse cloud with a mean radius k_{rms} given by

$$(\text{mean energy per particle}) = \frac{3}{2}kT = \frac{\hbar^2 k_{\text{rms}}^2}{2m} \quad (15.26)$$

Each particle has many unoccupied states to choose from, and particles seldom vie for the same single-particle state — the same point in k space. Consequently, the particle type, fermion or boson, is largely irrelevant. Ordinary gases, such as air at STP, are in this classical regime.



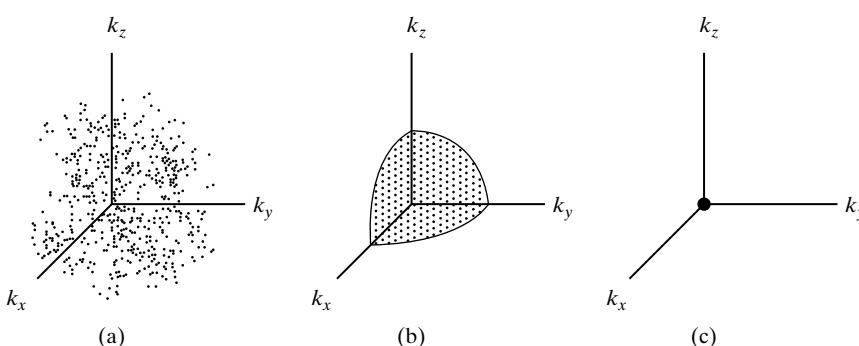
516 Chapter 15 • Statistical Mechanics

FIGURE 15.12

The k -space picture of an ideal gas in three different regimes. Each dot represents the state of one particle, given by the three numbers k_x , k_y , and k_z , all of which are positive.

(a) The classical, nondegenerate regime occurring at high temperature and low density.

(b) A degenerate gas of fermions at low temperature and high density. The N fermions occupy the N states of lowest energy, that is the N allowed points closest to the origin. (c) A degenerate gas of bosons, also at very low temperature and high density. All N bosons occupy the single state of lowest energy.



If we now lower the temperature so that the average energy per particle drops, then occupied states begin to crowd near the origin of k space, and particles compete for a limited number of k -space points. As discussed in Section 13.11, in this low-temperature, *degenerate quantum gas limit*, the behavior of the system depends critically on the type of particle involved. Bosons will happily share occupancy of a k -space point with an unlimited number of other particles. In contrast, fermions obey the Pauli exclusion principle and allow no more than one particle per k -space point if we neglect the particles' spin. If the particles have spin half, then two particles are allowed at each k -space point, one spin up and one spin down.

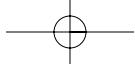
If the particles are fermions, then the N particles will fill the N lowest energy states nearest the origin. The filled states occupy an eighth of a sphere in k space, as in Figure 15.12(b). In this case the average energy per particle is very roughly equal to the Fermi energy, the energy of the highest occupied states, and is independent of the temperature. (More exactly, at $T = 0$, the average energy per particle and the Fermi energy are related by $\langle E \rangle = (3/5)E_F$; see Problem 15.39.) Recall from Chapter 13 that examples of degenerate Fermi gases include electrons in a metal, electrons in a white dwarf star, and neutrons in a neutron star.

If the particles are bosons, then as the absolute temperature is lowered to zero, all particles crowd into the lowest energy single-particle state, the state nearest the origin, as indicated in Figure 15.12(c). This many-particle system is then a Bose-Einstein condensate, as described in Section 13.12. Such an exotic state requires extremely low temperatures and is not normally encountered in nature; examples of degenerate boson gases include superconductors and superfluids. For the remainder of this chapter, we will consider only ordinary gases in the classical nondegenerate regime.

15.8 Energy and Speed Distributions in an Ideal Gas

In this section we apply the Boltzmann relation (15.6) to a *single* particle in the 3-D ideal gas described in the previous section. Because all particles in a gas are statistically identical in behavior, any one particle is representative of the behavior of all particles. We will derive expressions for the distribution of energies and the distribution of speeds of the particles of a gas in the classical, nondegenerate regime.

The Boltzmann relation in the form (15.6) applies to any quantum system with a discrete spectrum of energies. A particle in a 3-D quantum ideal gas has a discrete energy spectrum; however, for a large container of gas, the energy states are so numerous and so closely spaced that it is appropriate to view

Section 15.8 • Energy and Speed Distributions in an Ideal Gas **517**

the states as a continuum. In our k -space picture of the gas (Figure 15.11), the single-particle quantum states are a distance $\Delta k = \pi/a$ apart, where a is the edge length of the cubical container. If the box is macroscopic, then on an atomic scale the edge length a is very large and $\Delta k = \pi/a$ is very small. There are many many quantum states densely packed in k space. Recalling that the energy of a state is proportional to the square of the k -space distance to the origin, we see that there is a nearly continuous distribution of energies.

When considering a system with a continuum of allowed energies, instead of asking what is the probability $P(E)$ that the system has a specific energy E , we should rather ask what is the probability that the system has some energy in a small range from E to $E + dE$. The probability *distribution* $p(E)$ is defined by the relation

$$p(E) dE = (\text{probability system has energy in the range } E \text{ to } E + dE)$$

Similarly, in place of the degeneracy $g(E)$, we use a **density of states** $D(E)$ defined by the relation

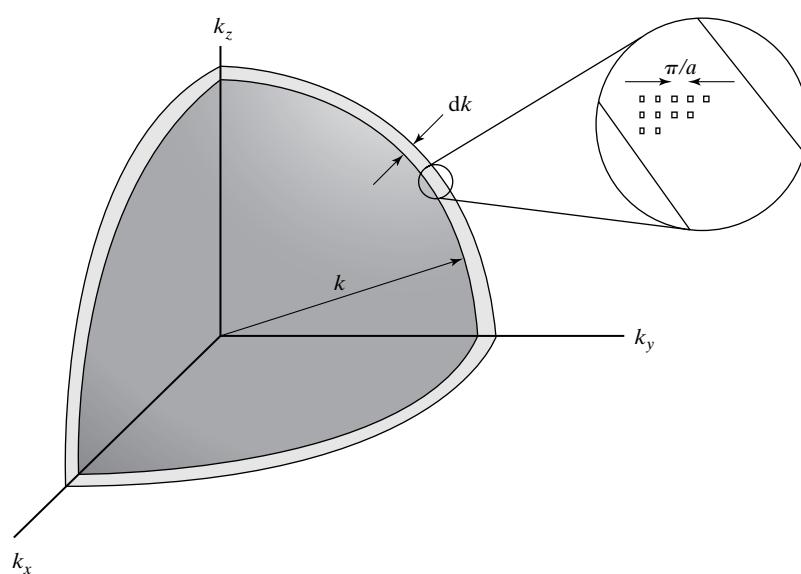
$$D(E) dE = \text{number of states with energies in the range } E \text{ to } E + dE$$

The Boltzmann relation (15.6), when applied to a system with a continuum of energies, then becomes

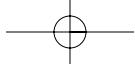
$$p(E) dE = \frac{D(E) e^{-E/kT} dE}{\int_0^\infty D(E) e^{-E/kT} dE} \quad (15.27)$$

As in (15.3), the normalization condition $\int_0^\infty p(E) dE = 1$ produces the constant $\int_0^\infty D(E) e^{-E/kT} dE$ in the denominator of (15.27).

We now compute the density of states $D(E)$ for a particle in a 3-D quantum ideal gas. Our strategy is first to compute how many states have $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ in a range k to $k + dk$, and then to change variables from k to E . In k space we consider a shell of radius k and thickness dk , as shown in Figure 15.13. Since the states within this shell are all nearly the same

**FIGURE 15.13**

A shell of nearly constant energy in k space.



518 Chapter 15 • Statistical Mechanics

distance from the origin, these states have nearly the same energy. As described in Section 15.7, the states within the shell form a cubic array of points. These points are separated by a distance $\Delta k = \pi/a$, so that corresponding to each point is a k -space volume $\Delta k^3 = (\pi/a)^3 = \pi^3/V$, where $V = a^3$ is the volume of the box containing the gas. The number of single-particle quantum states within this shell is given by

$$\begin{aligned} \left(\begin{array}{c} \text{number of states in} \\ \text{range } k \text{ to } k + dk \end{array} \right) &= \frac{\text{volume of the shell}}{\text{volume per } k\text{-space point}} \\ &= \frac{(1/8)4\pi k^2 dk}{(\pi/a)^3} = \frac{V}{2\pi^2} k^2 dk \quad (15.28) \end{aligned}$$

The factor of $(1/8)$ in the volume of the shell arises because only one octant of the full sphere appears in k space, as in Figure 15.13.

We now change variables from k to energy E . Corresponding to the wave number k is an energy $E = \hbar^2 k^2 / 2m$. Corresponding to the k increment dk is an energy increment

$$dE = \frac{dE}{dk} dk = \frac{\hbar^2 k}{m} dk$$

We substitute $dk = (m/\hbar^2 k) dE$ and $k = \sqrt{2mE}/\hbar$ into (15.28) and note that the number of states in the range k to $k + dk$ is the same as the number of states in the corresponding energy interval E to $E + dE$, which is $D(E) dE$. Therefore, (15.28) becomes

$$D(E) dE = \frac{V}{2\pi^2} \sqrt{2} \frac{m^{3/2}}{\hbar^3} \sqrt{E} dE \quad (15.29)$$

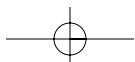
Inserting this expression for $D(E) dE$ into (15.27) yields the energy probability distribution

$$p(E) dE = \frac{\sqrt{E} e^{-E/kT} dE}{\int_0^\infty \sqrt{E} e^{-E/kT} dE} \quad (15.30)$$

Notice that the messy constant $(V/\sqrt{2}\pi^2)m^{3/2}\hbar^{-3}$ appearing in (15.29) cancels out in (15.30). The integral in the denominator of (15.30) is given in Appendix B, leading to our final expression for the probability distribution of single-particle energies in an ideal gas,

$$p(E) dE = \frac{2}{\sqrt{\pi}(kT)^{3/2}} \sqrt{E} e^{-E/kT} dE \quad (15.31)$$

This distribution is plotted in Figure 15.14. We see that a small fraction of the gas molecules have energies near zero, corresponding to molecules that are momentarily near rest. The peak in the distribution occurs at roughly energy $E = kT$. There is a long tail out to high energies, but very few molecules have energies greater than several kT .



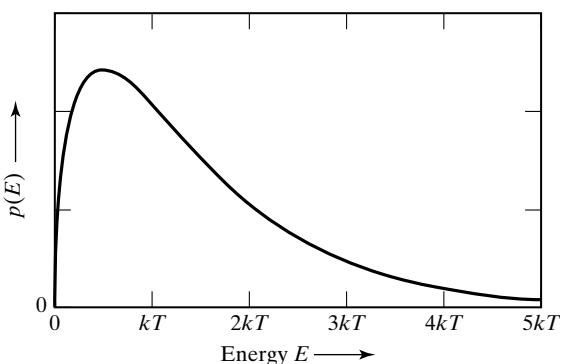


FIGURE 15.14
Distribution of single-particle energies in an ideal gas.

Computing the distribution of speeds of the gas molecules from the distribution of energies is mathematically straightforward, as we show below, but is conceptually subtle. Shifting from a description of energies to a description of speeds requires a delicate mixing of quantum and classical language. We have been describing a *quantum* ideal gas whose single-particle states are described by wave functions extending over the entire volume of the container. These wave functions are stationary states, standing wave states with a definite energy E and a definite wavelength λ . How do we reconcile our quantum-mechanical view of stationary, delocalized states with our classical view of localized particles moving with some velocity? The key to reconciling these apparently contradictory descriptions is wave packets.

As described in Section 6.7, wave-packet states have a spread in wavelengths $\Delta\lambda$ and a corresponding spread in momenta $\Delta p = \Delta(\hbar/\lambda)$, as well as a spread in angular frequencies $\Delta\omega$ and a corresponding spread in energies $\Delta E = \hbar\Delta\omega$. Although the wave packets do not have a single definite energy (as the stationary states do), they do have a well-defined average energy E . And though they do not possess a single definite momentum, they do have an average momentum p and a corresponding average speed $v = p/m$. As shown in Section 6.10, the average energy E and average speed v of the wave packet are related in the usual classical manner $E = \frac{1}{2}mv^2$.

Without further ado, let us write $E = \frac{1}{2}mv^2$ and perform a change of variables from E to v to transform the energy distribution (15.31) into a distribution of speeds. Corresponding to every energy E is a speed $v = \sqrt{2E/m}$ and corresponding to $p(E) dE$, the probability of finding a particle in a state with some energy between E and $E + dE$, is $p(v) dv$, the probability of finding the particle with corresponding speed between v and $v + dv$. Furthermore, since $E = \frac{1}{2}mv^2$,

$$dE = \frac{dE}{dv} dv = \frac{d}{dv} \left(\frac{1}{2}mv^2 \right) dv = mv dv$$

Substituting for E , \sqrt{E} , and dE in (15.31) yields

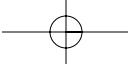
$$p(v) dv = \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT} \right)^{3/2} v^2 e^{-mv^2/2kT} dv \quad (15.32)$$

This expression is called the **Maxwell speed distribution** because it was first derived by Maxwell, using an entirely different approach before the development of either quantum mechanics or statistical mechanics.

James Clerk Maxwell
(1831–1879, Scottish)



A child prodigy with wide interests in math and physics, young Maxwell published important papers on geometry, the rings of Saturn, elastic solids, and the perception of color. Shy, quiet, and unassuming, he amazed colleagues with the breadth of his knowledge. Though a brilliant lecturer, he was once turned down for a university position because he lacked “the power of oral exposition proceeding on the supposition of imperfect knowledge or even total ignorance on the part of pupils.” Maxwell’s theory of electromagnetism, which describes light as an electromagnetic wave, is probably the greatest accomplishment of nineteenth-century physics. His pioneering work in the kinetic theory of gases helped establish statistical mechanics and the atomic hypothesis.



520 Chapter 15 • Statistical Mechanics

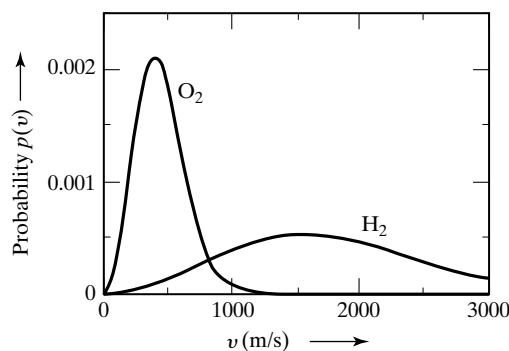


FIGURE 15.15

Speed distribution of oxygen and hydrogen gas at $T = 300\text{ K}$.

The Maxwell speed distribution (15.32) is plotted in Figure 15.15 for both O₂ molecules and lighter, faster H₂ molecules. The distribution of speeds is seen to be qualitatively similar to the energy distribution. There is a peak near the speed v at which $\frac{1}{2}mv^2 = \frac{3}{2}kT$. There is a tail out to higher speeds, but few molecules have speeds higher than about three times the most probable speed, which would correspond to an energy nine times the most probable energy. This figure explains why there is almost no hydrogen in the earth's atmosphere. For hydrogen, the tail of the speed distribution extends beyond the earth's escape speed (about 10 km/s) just enough so that any hydrogen in the atmosphere eventually leaks into space. Smaller astronomical bodies, such as the moon and asteroids, have such low escape velocities that all gases on their surfaces escape to space.

The Maxwell speed distribution has been experimentally verified with apparatus of the type shown in Fig. 15.16. An oven, containing a known gas at a well-controlled temperature, has a small hole through which gas molecules can escape into an evacuated test chamber. For a given rotation rate of the spinning cylinder, only molecules of a particular speed can pass through the slot of the spinning cylinder and strike the detector. The speed distribution is determined by measuring the detector signal as a function of the rotation rate of the cylinder.

Given the probability distributions $p(E)$ and $p(v)$, we can compute the average energy or mean energy $\langle E \rangle$ and mean speed $\langle v \rangle$ of a particle by using the expectation-value formula (7.69),

$$\langle E \rangle = \int_0^\infty Ep(E) dE$$

and

$$\langle v \rangle = \int_0^\infty vp(v) dv$$

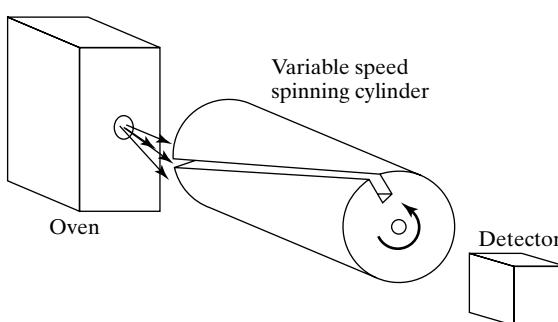


FIGURE 15.16

Apparatus to measure the speed distribution of gas molecules.

These are often called the “thermal average” values. We leave it as exercises (Problems 15.47 and 15.48) to show that the thermal average values of E and v are

$$\langle E \rangle = \int_0^\infty Ep(E) dE = \frac{3}{2}kT$$

and

$$\langle v \rangle = \int_0^\infty vp(v) dv = \sqrt{\frac{8kT}{\pi m}}$$

Of course, we already knew the first result, $\langle E \rangle = \frac{3}{2}kT$, from the equipartition theorem.

It is not hard to show that for an ideal gas, the rms speed $v_{\text{rms}} = \sqrt{\langle v^2 \rangle}$ is given by $v_{\text{rms}} = \sqrt{3kT/m}$, which is slightly larger than the mean speed $\langle v \rangle$. The rms average speed is often a more useful quantity than the mean speed because the rms speed is more clearly related to the mean energy via the equation $\langle E \rangle = \frac{1}{2}m\langle v^2 \rangle$.

15.9 Heat Capacities

The energy levels of atoms or molecules in the gaseous state can be readily determined by measuring the spectrum of light emitted when the gas is heated. Each line in the spectrum corresponds to a transition between a pair of discrete energy levels in the atom or molecule. In contrast, it is quite difficult to measure directly the energy spectrum of many-particle condensed-matter systems such a block of copper or a container of liquid helium — systems in which the energy levels are so closely spaced that they form a continuum. However, important clues about the energy spectrum can be gleaned from a measurement of the system’s heat capacity. The heat capacity C of a system is defined as

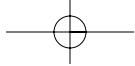
$$C = \frac{d\langle E \rangle}{dT} \quad (15.33)$$

where $\langle E \rangle$ is the thermal average energy of the system. To raise the temperature of a system by a small amount δT , we must input an amount of energy $\delta E = C \delta T$. Usually the energy is added to the system in the form of heat, so the heat capacity C is the amount of heat absorbed (or released) when the temperature is raised (or lowered) by one degree Kelvin. As we will see, the dependence of the heat capacity on temperature yields information on the spectrum of quantum energy levels of the system.

The heat capacity of a monatomic ideal gas has a very simple form. According to the equipartition theorem (15.1), the energy per degree of freedom of a gas is $\frac{1}{2}kT$. In a gas, the number of degrees of freedom per atom is 3 (one each for translational motion in the x , y , and z directions). If the total number of atoms is N , the total energy $\langle E \rangle$, which we write as* U is

$$\begin{aligned} U &= (\text{number of degrees of freedom}) \times (\frac{1}{2}kT \text{ per degree of freedom}) \\ &= (3N)(\frac{1}{2}kT) = \frac{3}{2}NkT \end{aligned}$$

*As is traditional in chemistry and statistical mechanics, we use the symbol U to represent the total internal energy, which in this case is the same as $\langle E \rangle$. Be careful not confuse this with our previous use of U for potential energy.



522 Chapter 15 • Statistical Mechanics

The heat capacity of a monatomic ideal gas is then

$$C = \frac{dU}{dT} = \frac{3}{2}Nk = \frac{3}{2}nR$$

where R is the universal gas constant $R = N_A k = 8.31 \text{ J/mole}\cdot\text{K}$ and $n = N/N_A$ is the number of moles. This temperature-independent value $\frac{3}{2}R$ per mole for the heat capacity of an ideal gas has been abundantly confirmed by experiment.

The heat capacity of a solid is more complicated than that of a gas. A solid can store energy in more ways than a gas can — it has more degrees of freedom — and, as we will see, the heat capacity of a solid is temperature-dependent. A crystalline solid with its springlike chemical bonds between atoms can store energy in the vibrations of the atomic lattice, leading to a “lattice heat capacity.” In addition, if the solid is a metal, the conduction electrons can store energy and contribute to the total heat capacity. This electronic heat capacity is absent in electrical insulators, and even in metals it is generally small compared to the lattice heat capacity. Here, we will consider insulating solids only.

An insulating solid with N atoms has $6N$ degrees of freedom. Recall (Section 15.2) that in a solid, each atom has six degrees of freedom because the total energy of each atom has three kinetic energy terms ($\frac{1}{2}mv_x^2, \frac{1}{2}mv_y^2, \frac{1}{2}mv_z^2$) and three potential energy terms ($\frac{1}{2}kx^2, \frac{1}{2}ky^2, \frac{1}{2}kz^2$). The equipartition theorem then leads us to predict that for *any* solid with N atoms, the total energy and lattice heat capacity are given by

$$U = (6N)\left(\frac{1}{2}kT\right) = 3NkT$$

and

$$C = \frac{dU}{dT} = 3Nk = 3nR$$

This result for the heat capacity of solids is called the **Dulong–Petit law**. It is found to agree with experiment at high temperatures, but at low temperatures, typically somewhat below room temperature, the heat capacity falls rapidly as the temperature is lowered and approaches zero as $T \rightarrow 0 \text{ K}$, as shown in Fig 15.17.

This disagreement between theory and experiment was deeply puzzling to physicists of the early twentieth century. It was Einstein who showed that the discrepancy is due to a quantum effect that causes a breakdown of the classical

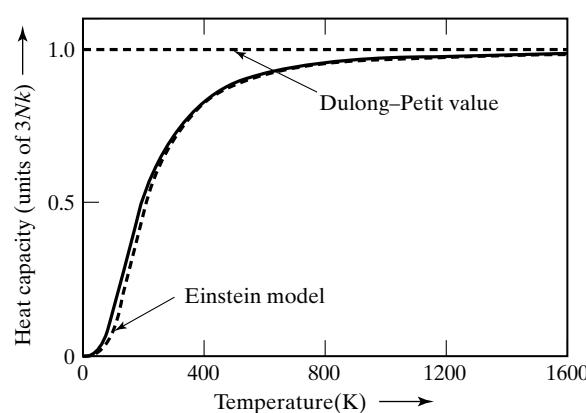


FIGURE 15.17

Measured heat capacity of germanium (solid line) and the predictions of the Einstein model and Dulong–Petit.

equipartition theorem. Recall that the equipartition theorem assumes that the energy of the system is continuous, not discrete. Real physical systems obey quantum mechanics and have quantized energies. If the temperature is high enough that $kT \gg \Delta E$, where ΔE is the separation between adjacent energy levels, then the energy spectrum can be approximated as a continuum and the equipartition theorem applies. But when the temperature is so low that $kT < \Delta E$, the equipartition theorem does not apply and the energy quantization has a profound effect on the heat capacity. The vanishing of the heat capacity at low temperatures in quantum systems can be understood qualitatively as follows: If the temperature is so low that $kT \ll \Delta E$, where ΔE is the energy gap between the ground state and the first excited state, then the system will almost always be in its ground state. A small increase in temperature will not give the system enough thermal energy to jump up to the first excited state and it will remain in the ground state. The excited states are said to be “frozen out.” There will be (almost) no increase in the mean energy $\langle E \rangle$ when the temperature is increased, so $C = d\langle E \rangle/dT \approx 0$.

To explain in detail the low-temperature heat capacity of solids (Fig. 15.17), we need a quantum calculation of the thermal average energy $U = \langle E \rangle$ of an atom in a solid. This calculation was first performed by Einstein.

The Einstein Model of Heat Capacity

An atom in a solid can be modeled as a simple 3-D harmonic oscillator — a mass m with potential energy $U = \frac{1}{2}kr^2$, where r is the distance from the equilibrium position and the spring constant k depends on the strength of the chemical bonds with neighboring atoms. We will begin by computing the heat capacity of a 1-D harmonic oscillator and then will generalize to 3-D. As described in Section 7.9, the energy levels of a 1-D harmonic oscillator are given by $E_n = (n + \frac{1}{2})\hbar\omega$, where $n = 0, 1, 2, \dots$ and $\omega = \sqrt{k/m}$. To simplify things, we can reset the zero of energy so that $E_n = n\varepsilon$, where $\varepsilon = \hbar\omega$. Using the Boltzmann relation (15.3), we have (all sums shown are over $n = 1, 2, \dots$)

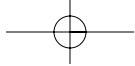
$$\begin{aligned}\langle E \rangle &= \frac{\sum E_n e^{-E_n/kT}}{\sum e^{-E_n/kT}} = \frac{\sum n\varepsilon e^{-n\varepsilon/kT}}{\sum e^{-n\varepsilon/kT}} \\ &= \frac{\varepsilon e^{-\varepsilon/kT}[1 + 2(e^{-\varepsilon/kT}) + 3(e^{-\varepsilon/kT})^2 + \dots]}{1 + e^{-\varepsilon/kT} + (e^{-\varepsilon/kT})^2 + \dots}\end{aligned}$$

The last expression can be simplified by making use of the following identities

$$\frac{1}{(1-x)} = 1 + x + x^2 + x^3 + \dots \quad \text{and} \quad \frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \dots \quad (15.34)$$

(See Problem 15.51.) The mean energy is then

$$\langle E \rangle = \frac{\varepsilon e^{-\varepsilon/kT}}{1 - e^{-\varepsilon/kT}} = \frac{\varepsilon}{e^{+\varepsilon/kT} - 1} \quad (15.35)$$



524 Chapter 15 • Statistical Mechanics

And the heat capacity of a single 1-D harmonic oscillator is

$$C_{1D} = \frac{d\langle E \rangle}{dT} = k \left(\frac{\varepsilon}{kT} \right)^2 \frac{e^{\varepsilon/kT}}{(e^{\varepsilon/kT} - 1)^2} \quad (15.36)$$

where $\varepsilon = \hbar\omega$. Notice that Boltzmann's constant k has the units of heat capacity (recall $E = kT$, and hence $k = E/T$), so (15.36) is written as k multiplied by dimensionless quantities.

A 3-D simple harmonic oscillator can be regarded as three independent 1-D harmonic oscillators (one each for the x , y , and z directions). The total energy of the 3-D oscillator is just the sum of the energies of the three 1-D oscillators, and we conclude that the heat capacity of a 3-D harmonic oscillator, which we use to model a single atom in a solid, is three times the expression in (15.36), $C_{\text{atom}} = 3C_{1D}$.

Einstein made the simplifying assumption that the atoms in a solid could be treated as *independent* 3-D harmonic oscillators, so that the total energy of the collection of N atoms is just the sum of the energies of the N independent atoms, $\langle E \rangle_{\text{solid}} = N\langle E \rangle_{\text{atom}}$ and the total heat capacity of the solid is simply

$$C_{\text{solid}} = \frac{d}{dT} \langle E \rangle_{\text{solid}} = NC_{\text{atom}} = 3NC_{1D}$$

There is an unknown parameter in this model, namely the value of $\varepsilon = \hbar\omega$, the energy-level spacing of the simple harmonic oscillator used to approximate the vibrational behavior of an atom. The parameter ε is regarded as a *fitting parameter*, and its value is adjusted to give the best agreement between theory and experiment. The resulting **Einstein model of heat capacity** is plotted in Fig. 15.17 along with experimental data on the heat capacity of germanium. We see that the model gives a very good fit with a temperature-independent heat capacity at high temperatures ($T > \hbar\omega$), and a vanishing heat capacity at low temperatures. At very low temperatures, ($0 < T \ll \hbar\omega$), the Einstein model does not give accurate quantitative results, because of a failure of the independent-oscillator approximation; nevertheless, the qualitative success of the model in explaining the heat capacity of solids was one of the earliest successes of quantum theory.

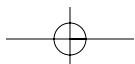
Example 15.7

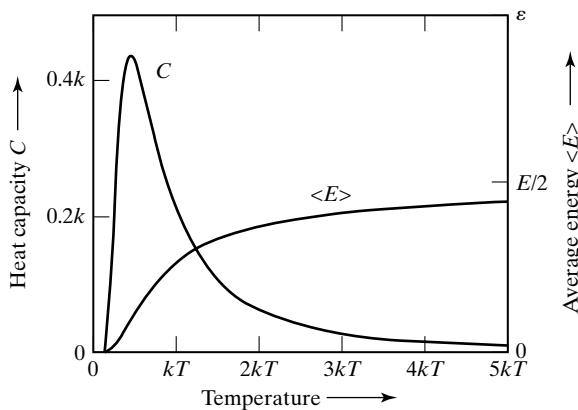
Consider again the two-level system described in Example 15.5. Compute the heat capacity C of the system as a function of temperature.

A two-level system, such as a spin magnetic moment in an external magnetic field, has a ground state with energy $E = 0$ and a single excited state with energy $E = \varepsilon$. The mean thermal energy of such a system at temperature T is

$$\langle E \rangle = \frac{\sum E_i e^{-E_i/kT}}{\sum e^{-E_i/kT}} = \frac{0 + \varepsilon e^{-\varepsilon/kT}}{1 + e^{-\varepsilon/kT}} = \frac{\varepsilon}{e^{\varepsilon/kT} + 1} \quad (15.37)$$

Both sums are over two states only: $i = 1, 2$.



**FIGURE 15.18**

Heat capacity and average energy of a two-level system.

The heat capacity is then

$$C = \frac{d\langle E \rangle}{dT} = k \frac{(\varepsilon/kT)^2 e^{\varepsilon/kT}}{(1 + e^{\varepsilon/kT})^2} \quad (15.38)$$

Notice the similarity between these expressions and the corresponding expressions (15.35) and (15.36) for a harmonic oscillator. (But notice that the minus sign in the harmonic-oscillator formulas has been replaced by a plus sign for the two-level system.) The heat capacity (15.38) is plotted in Fig. 15.18 along with the mean energy (15.37). Notice that the peak in the heat capacity $C = d\langle E \rangle/dT$ occurs where the slope of $\langle E \rangle$ versus T is maximum.

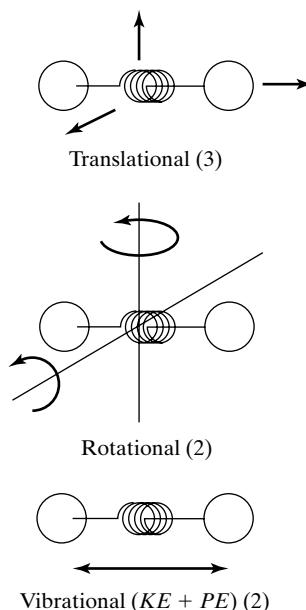
Diatom Gases

At the beginning of this section, we argued that the heat capacity of an ideal monatomic gas is $\frac{3}{2}Nk$, where the factor of $\frac{3}{2}$ came from the fact that each of the three degrees of freedom (motion in the x , y , and z directions) of a gas atom contributes $\frac{1}{2}kT$ to the energy. It may come as a surprise then to learn that at room temperature the heat capacity of oxygen and nitrogen, the principle components of air, is $\frac{5}{2}Nk$, not $\frac{3}{2}Nk$. (When dealing with molecular gases, N is the number of molecules in the gas, not the number of atoms.) The reason is that molecules with two or more atoms have **internal degrees of freedom**. Oxygen and nitrogen, as well as hydrogen, chlorine, and many other gases, form *diatomic* molecules, and these multi-atom molecules can rotate and vibrate, as well as translate. Rotation contributes two degrees of freedom, since a dumbbell-shaped diatomic molecule can rotate about two orthogonal axes (Fig. 15.19). Vibration contributes two more degrees of freedom, radial kinetic energy and elastic potential energy, similar to a simple 1-D mass on a spring.

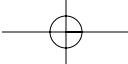
So, in addition to the three translational degrees of freedom, we have four internal degrees of freedom, for a total of seven degrees of freedom. The corresponding seven quadratic terms in the energy are

$$U = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2 + \frac{1}{2}I\omega_x^2 + \frac{1}{2}I\omega_y^2 + \frac{1}{2}\mu(dr/dt)^2 + \frac{1}{2}k(r - r_0)^2$$

(Can you identify each term and the meaning of each symbol?) With seven degrees of freedom, why then is the heat capacity not $(7/2)Nk$? At sufficiently high temperature, it is! The heat capacity of diatomic molecules is temperature-dependent. Recall that a degree of freedom, which is often called a “mode,”

**FIGURE 15.19**

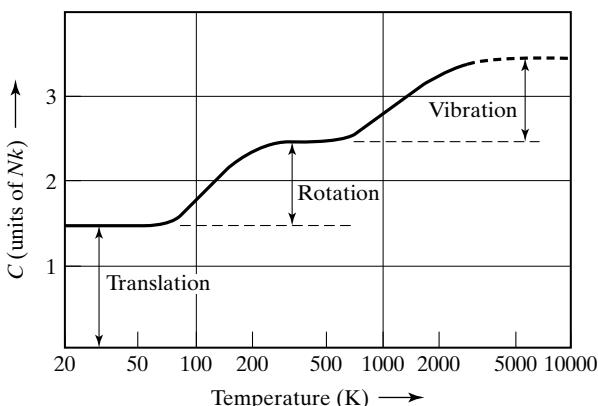
The seven degrees of freedom of a diatomic molecule.



526 Chapter 15 • Statistical Mechanics

FIGURE 15.20

Heat capacity of hydrogen, a diatomic gas as a function of temperature. The hydrogen molecule dissociates at high temperatures, before the vibrational modes are fully excited.



fully contributes to the heat capacity only when $kT \gg \Delta E$, where ΔE is the energy-level spacing for the quantum levels associated with that mode. At very low temperatures ($T \lesssim 100$ K) both the rotational and vibrational degrees of freedom are frozen out, and the heat capacity is $\frac{3}{2}Nk$, since only the center-of-mass kinetic energy contributes to the heat capacity. We saw in Chapter 12 that the energy-level spacing for the rotational modes of molecules are smaller than the level spacing for the vibrational modes. At intermediate temperatures — around room temperature — the two rotational degrees of freedom become thermally activated, but the vibrational degrees are still frozen out, leading to a heat capacity of $\frac{5}{2}Nk$. At sufficiently high temperatures, kT is greater than the quantum of excitation of the vibrational modes and all degrees of freedom contribute to the heat capacity $\frac{7}{2}Nk$. Fig. 15.20 shows the temperature dependence of the heat capacity of hydrogen gas, which clearly exhibits these three regimes. We leave it as exercises (Problems 15.61 and 15.62) to estimate the temperatures at which the vibrational and rotational degrees of freedom are activated.

CHECKLIST FOR CHAPTER 15

CONCEPT	DETAILS
Microstates and macrostates	Complete specification of all microscopic variables versus specification of statistical averages of a few macroscopic variables. (Section 15.1)
Degrees of freedom	Number of quadratic terms in detailed expression of total energy. (Section 15.2)
Equipartition theorem	The total energy of a classical system is shared equally among its degrees of freedom. (Section 15.2)
Temperature	Amount of atomic “jiggling.” $\frac{1}{2}kT = \text{energy per degree of freedom in a classical system}$ (15.1)
Boltzmann relation: probability that a system is in a particular quantum state i	$P(\text{state } i) = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}} \quad (15.3)$
Probability that a system has an energy E	$P(E) = g(E)P(i) \quad (15.6)$
Equal-probability hypothesis	All accessible microstates are equally likely to occur. (Section 15.4)

Counting microstates	The probability of a situation = $\frac{(\text{number of ways that situation can occur})}{(\text{total number of possible situations})}$ (Section 15.4)
Number of ways of choosing N indistinguishable objects from among T objects	$g = \frac{T!}{(T - N)!N!} \quad (15.7)$
Origin of the Boltzmann relation*	For an isolated (system + reservoir), if the system is in a state i with more energy, then the reservoir has less energy and therefore fewer microstates. (Section 15.5)
Entropy and the second law*	The entropy of a closed system cannot decrease. (Section 15.6)
thermodynamic definition of entropy	$dS = dQ/T \quad (15.17)$
statistical definition of entropy	$S = k \ln g \quad (15.18)$
Quantum ideal gas	Each one-particle state is specified by a point in k space. (Section 15.7)
Density of states $D(E)$ of ideal gas	$D(E) dE = \text{number of states with energies between } E \text{ and } E + dE, D(E) \propto \sqrt{E} \quad (15.29)$
Maxwell speed distribution $p(v)$	$p(v) dv = \text{probability that particle has speed between } v \text{ and } v + dv \quad (15.32)$
Heat capacity	$C = d\langle E \rangle / dT \quad (15.33)$
Dulong–Petit law	For a classical solid, energy $U = 3NkT$, so $C = dU/dT = 3Nk$ (Section 15.9) When $kT \ll \Delta E$, the excited modes of a quantum solid are frozen out, leading to a vanishingly small heat capacity. (Section 15.9)
Einstein model of heat capacity	In addition to translational modes, the rotational and vibrational modes can contribute to the heat capacity. (Section 15.9)
diatomic gases	

PROBLEMS FOR CHAPTER 15

SECTION 15.1 (Introduction)

- 15.1** • Neon is a noble element and therefore forms a gas of monatomic molecules at all but the lowest temperatures. However, neon freezes and forms a crystalline solid at temperatures below $T = 26$ K. Classically, how many degrees of freedom are in (a) a mole of neon gas at room temperature? (b) A mole of solid neon at $T = 10$ K?

SECTION 15.2 (Temperature)

- 15.2** • What is the temperature in kelvins that corresponds to an energy of kT equal to (a) 0.001 eV (b) 0.1 eV (c) 10 eV (d) 1000 eV?
- 15.3** • (a) What is the rms speed of a helium atom in a sample of pure helium gas at room temperature ($T \approx 293$ K)? (b) How do you expect the speed of sound in helium gas to compare with that in air at the same temperature?
- 15.4** • Radon is a heavy inert gas, with atomic number $Z = 86$. What is the rms speed of radon atoms in air at room temperature ($T \approx 293$ K)? Is the rms speed of the atoms in pure radon gas greater than, less than, or the same as that of radon atoms mixed in air?
- 15.5** • A water droplet in a cloud is typically 5 microns in diameter (1 micron = $1 \mu\text{m} = 10^{-6}$ m). What is the rms speed of such a droplet at room temperature ($T \approx 293$ K)?
- 15.6** • (a) What is the ratio of the kinetic energies of O_2 molecules to that of N_2 molecules in air at room temperature? (b) What is the ratio of the rms speed of O_2 molecules to that of N_2 molecules in air at room temperature? (c) Are these ratios temperature-dependent?
- 15.7** •• (a) What is the total energy E_{tot} of a mole of helium gas at room temperature, $T \approx 293$ K? [Hint: This is just $N\langle K \rangle$, where N is the number of atoms and $\langle K \rangle$ is the average kinetic energy per atom.] (b) By how many kelvins would this energy raise the temperature of a cup of water (200 grams, say)?
- 15.8** •• (a) Write an expression for the total energy E_{tot} of a mole of a monatomic gas at temperature T . (See Problem 15.7.) (b) Use this to find the (constant-

528 Chapter 15 • Statistical Mechanics

volume) molar specific heat C_v of the gas. This is the energy needed to raise the temperature of 1 mol by 1 K, so is equal to dE_{tot}/dT .

- 15.9 •••** In his *Feynman Lectures in Physics* (Vol. 1), Richard Feynman gives the following proof of the equipartition theorem for the special case of molecules in an ideal gas: Consider two molecules of masses m_1 and m_2 with initial velocity vectors v_1 and v_2 , prior to collision. The velocity of the center of mass \mathbf{v}_{cm} is defined by the equation $(m_1 + m_2)\mathbf{v}_{\text{cm}} = (m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$.

The relative velocity \mathbf{w} of molecules is defined as $\mathbf{w} = (\mathbf{v}_1 - \mathbf{v}_2)$. One can argue that if the system is in thermal equilibrium, the direction of the center-of-mass motion is completely uncorrelated to the direction of the relative velocity, so that $\langle \mathbf{w} \cdot \mathbf{v}_{\text{cm}} \rangle = 0$, where the brackets $\langle \dots \rangle$ represents an average over all pairs of molecules. (Since the directions are random, the dot-product is positive as often as negative and it averages to zero.) Use this relation to argue that

$$\frac{1}{2}m_1\langle v_1^2 \rangle = \frac{1}{2}m_2\langle v_2^2 \rangle$$

SECTION 15.3 (The Boltzmann Factor)

- 15.10 •** Consider the quantum-mechanical system consisting of a particle of mass m in a one-dimensional rigid box of length a , (often called an infinite square well), as described in Section 7.4. Assume that the system is in thermal equilibrium at temperature T . **(a)** Write an expression for the ratio $P(E_2)/P(E_1)$, that is, the ratio of the probability that the system is in the first excited state to that in the ground state. **(b)** What is the approximate temperature below which the system is unlikely to be found in any of its excited states?

- 15.11 •** Consider a gas of hydrogen atoms at a temperature T . At what temperature is the ratio of the number of atoms in the first excited state to the number of atoms in the ground state equal to 1/10; that is, at what temperature is $P(E_2)/P(E_1) = 1/10$? [Don't forget that these states have degeneracies.]

- 15.12 •• Physical interpretation of the partition function.** Consider a single-particle quantum system whose states are labeled with an index $i = 1, 2, 3, \dots$ with corresponding energies E_1, E_2, E_3, \dots Set the zero of energy at the ground-state energy so that $E_1 = 0$. Argue that if the absolute temperature T is such that $kT = E_n$, where n is some integer, then n is the approximate value of the partition function (15.4). Therefore, the partition function is equal to the number of states that are likely to be occupied.

- 15.13 ••** A pocket compass with a magnetized iron needle is at rest in the earth's magnetic field, $B \approx 10^{-4}$ T. **(a)** Given that the mass of the needle is 0.2 grams, estimate its magnetic moment. Assume that each atom in the needle has a magnetic moment of one Bohr magneton and that the moments of all the atoms are fully aligned. **(b)** Find the ratio of the probability that the needle is pointing along the earth's field to that of its pointing in the opposite direction. [It is

legitimate to use the Boltzmann relation in this classical situation.]

- 15.14 ••** Consider an electron in an external magnetic field of 1.0 T, and recall that the energy of an electron in a magnetic field is given by Eq. (9.18) as $E = m\mu_B B$, where $m = \pm 1/2$. **(a)** Make a plot of the ratio of the probability that the electron's moment is aligned with the field (the low-energy state) to the probability that it is anti-aligned (the high-energy state) versus the absolute temperature T . **(b)** What is the ratio at room temperature? **(c)** How low does the temperature have to be in order for this ratio to be 2?

- 15.15 ••** At what temperature would one be just as likely to find a hydrogen atom in its first excited level as in its ground level? [Don't forget about degeneracies.]

- 15.16 ••** As described in Section 7.9, a 1-D quantum-mechanical harmonic oscillator has a particularly simple spectrum of energies: $E_n = (n + \frac{1}{2})\hbar\omega$, where $n = 0, 1, 2, \dots$, ω is a constant, and all the states are nondegenerate. **(a)** At temperature T , what is the ratio $P(E_0)/P(E_1)$, that is, the ratio of the probability that the system is in the ground ($n = 0$) state to the probability that it is in the first excited ($n = 1$) state? **(b)** Sketch the ratio $P(E_0)/P(E_1)$ as a function of the parameter $kT/\hbar\omega$.

- 15.17 •••** Consider the 1-D harmonic oscillator of the previous problem. **(a)** Write down the partition function (15.4) for this system and sum the infinite series. [Remember that $1 + x + x^2 + x^3 + \dots = 1/(1 - x)$.] **(b)** Sketch the probabilities $P(E_0)$ and $P(E_1)$ as functions of T .

- 15.18 •••** Consider a quantum system consisting of two spin-half particles in a magnetic field B . This system has four possible states, which we can indicate schematically as (1) $\uparrow\uparrow$, (2) $\downarrow\downarrow$, (3) $\uparrow\downarrow$, and (4) $\downarrow\uparrow$. States 3 and 4 both have energy $E = 0$, state 2 has energy $E = +\epsilon$, and state 1 has energy $E = -\epsilon$, where $\epsilon = 2\mu_B B$. **(a)** Write down an expression for the partition function for this system. [Note that the $E = 0$ state has a degeneracy of two.] **(b)** Write expressions for the probabilities that the system is in each of the three energy levels $E = +\epsilon, 0, -\epsilon$, as a function of temperature. **(c)** Sketch these probabilities as a function of the parameter kT/ϵ .

SECTION 15.4 (Counting Microstates: The Equal-Probability Hypothesis)

- 15.19 • (a)** What is the probability that in a deal of 2 cards from a randomly shuffled deck you will get 2 hearts? **(b)** What is the probability of turning up 2 hearts if the first card is replaced and the deck shuffled before the second card is turned?

- 15.20 • (a)** What is the probability that in a roll of 5 dice you will get 5 sixes? **(b)** What is it that you will get no sixes?

- 15.21 •• (a)** What is the probability that, from a randomly shuffled deck, you will be dealt a hand of 5 cards that are all hearts. **(b)** What is the probability that you will be dealt 4 kings and an ace in that order (4 kings followed by an ace)?

15.22 •• A coin is flipped N times. The probability that the coin will come up heads all N times is $P(N) = 1/2^N$. **(a)** Show that this probability can also be written as $P(N) = 10^{-N \ln 2 / \ln 10}$. [Hint: Note that $x = e^{\ln x} = 10^{\log x}$, from which it follows that $\ln x = \ln[10^{\log x}] = (\log x)(\ln 10)$.] **(b)** For what value of N does the probability of N heads in a row fall below 10^{-8} ? (10^{-8} is about the probability of winning a large state lottery.)

15.23 •• **(a)** What is the probability that, from a randomly shuffled deck, you will be dealt the cards (A♦, A♥, A♠, A♣, K♣) *in that order*? **(b)** What is the probability that, from a randomly shuffled deck, you will be dealt the same 5 cards *in any order*?

15.24 •• For very large numbers, the factorial of N can be estimated by *Stirling's approximation* $N! \approx (\sqrt{2\pi N}) N^N e^{-N}$. **(a)** What is the fractional error in this approximation for $N = 25$ and $N = 60$? **(b)** Stirling's approximation can also be written as $\ln(N!) \approx \frac{1}{2} \ln(2\pi) + \frac{1}{2} \ln N + N \ln N - N$. Often, to simplify calculations, a further approximation is made and the formula is written as $\ln(N!) \approx N \ln N - N$. If $N = 10^{23}$, how do the values of these two versions of Stirling's approximation compare? What percent error is made in the computation of $\ln(N!)$ by using the shorter, more approximate version compared to the longer, more precise version (for $N = 25$ and 60 again)? [See also Problem 15.63.]

15.25 •• Show that Eq. (15.11)

$$P(\text{left side only}) = \frac{H!}{(2H)!} \frac{(2H - N)!}{(H - N)!}$$

reduces to the simpler equation (15.8) $P(N) = 1/2^N$ in the limit $H \gg N \gg 1$. [Hint: To compute the factorial of large numbers, use *Stirling's approximation*: $\ln N! \approx N \ln N - N$. Note also that if $H \gg N$, then $\ln(H - N) \approx \ln H$.]

15.26 ••• Consider N identical indistinguishable bosons in a sealed container, and assume that the particles can move freely between the left and right halves of the container. As in Section 15.4, model the quantum states of this system in the following simple manner: Assume that there are H different single-particle quantum states on the left half of the box, that is, states with wave functions localized on the left, and there are another H states on the right. Compute the probability that all N bosons will be found in the left half of the box. [Hint: See Problem 15.29.]

15.27 ••• Consider a room full of air molecules. Each molecule can be considered to be in one of two states: in the right half of the room or in the left half. The probability of each of these states is $1/2$. If there are N distinguishable molecules in the room, then the total number of ways of arranging the molecules is 2^N . **(a)** Argue that the number of ways of arranging the N distinguishable molecules with N_L molecules on the left is

$$P_L = \frac{N!}{(N - N_L)! N_L!}$$

(b) If there are 100 molecules in the room, what is the ratio $(P_L = 0.5)/(P_L = 0.6)$; that is, how much more likely is it that the molecules are evenly distributed compared to having 60 on the left and 40 on the right? **(c)** If there are 10^{25} molecules in the room, what is the ratio $(P_L = 0.500)/(P_L = 0.501)$? To compute the factorial of large numbers, use *Stirling's approximation*: $\ln N! \approx N \ln N - N$.

15.28 ••• Continuing with the situation in the previous problem, the number of molecules in the right and left halves of a room containing N molecules can be written as $N_R = (N/2)(1 + \Delta)$, $N_L = (N/2)(1 - \Delta)$. For N large, we expect the number of molecules in the two halves to be nearly equal so $\Delta \ll 1$. **(a)** Show that the probability of a fractional discrepancy of 2Δ more molecules on the right than on the left is given by

$$P(\Delta) = Ce^{-\Delta^2 N}$$

where C is a constant that does not depend on Δ . [Hint: You will need to use *Stirling's approximation*, $\ln N! \approx N \ln N - N$, and the relation $\ln(1 + x) \approx x$, valid for $x \ll 1$. Begin by writing $P = N!/(N_R! N_L!)$ and take the \ln of both sides.] **(b)** For a room containing 10^{25} molecules, what is the ratio $P(\Delta = 0.001)/P(\Delta = 0)$? **(c)** What values of Δ are likely to actually occur in a room?

SECTION 15.5 (The Origin of the Boltzmann Relation*)

15.29 •• Prove Eq. (15.14),

$$g(r) = \frac{(N + r - 1)!}{r!(N - 1)!}$$

where $g(r)$ is the number of ways of placing r indistinguishable objects in N distinct boxes. A particular arrangement of the objects among the boxes can be represented schematically by a series of r dots sprinkled among $N + 1$ lines, like so:



The figure above represents $r = 4$ objects placed in $N = 6$ boxes. Note that the two lines on the far right and far left must remain fixed on the outside positions, but the remaining $(N - 1)$ lines and r dots can occur in *any order*.

15.30 •• Show that Eq. (15.14),

$$g(r) = \frac{(N + r - 1)!}{r!(N - 1)!}$$

is approximately equal to r^N in the limit $r \gg N \gg 1$. [Hint: Use *Stirling's approximation*: $\ln N! \approx N \ln N - N$. You will need to make several approximations along the way, including, for instance, $\ln(N + r - 1) \approx \ln(r)$.]

15.31 ••• Show that Eq. (15.16) $P(s) \propto e^{-s/(m/N)}$ can be derived from Eq. (15.15),

$$P(s) \propto g_R(m - s) = \frac{[N + (m - s) - 1]!}{(m - s)!(N - 1)!}$$

530 Chapter 15 • Statistical Mechanics

in the limit of large N and $m \gg N \gg 1$ and $m \gg s$. [Hint: Use Stirling's approximation: $\ln N! \approx N \ln N - N$, valid for large N . Also, note that, if $x \ll 1$, then $\ln(1 + x) \approx x$.]

SECTION 15.6 (Entropy and the Second Law of Thermodynamics *)

15.32 • Consider Example 15.6 in which a gas is allowed to expand from a volume V_0 into a volume $2V_0$ when a barrier is broken. Using both of the methods given in that example, find the change in entropy if the gas is allowed to expand in this way from volume V_0 to $3V_0$.

15.33 • An ideal gas of N spinless monatomic atoms is at temperature T and volume V . What is the entropy increase of the gas when the temperature is doubled at constant volume? [Hint: Use a thermodynamic argument, not a statistical mechanical one, and remember that for a monatomic ideal gas, the average energy per atom is $\frac{3}{2}kT$.]

15.34 • A bar of ferromagnetic material consists of N atoms, each with spin half. The spin of each atom can point in one of two directions: up or down. When the bar is fully magnetized, all the spins point in the same directions, like so: $\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\cdots$; it is completely demagnetized when the spins are in random directions, like so: $\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\downarrow\cdots$. (a) How many different ways g can the N spins be arranged? (Note that these are N distinguishable spins since they are on a fixed lattice.) (b) What is the increase in entropy of the bar magnet when it is taken from a fully magnetized state to a fully demagnetized state? (c) Repeat part (b), but with a magnet made of spin $\frac{3}{2}$ atoms, which can exist in any one of four states $\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$.

15.35 ••• An ideal gas of N spinless monatomic atoms is at a temperature T and volume V . (a) What is the entropy increase of the gas when the volume of the gas is doubled isothermally (at constant temperature)? (b) What is the entropy change of the gas when the volume of the gas is doubled reversibly (slow, steady volume change) and adiabatically (no heat allowed into or out of the gas)?

15.36 ••• *Entropy of Mixing.* Consider a box containing N particles. A partition separates the right and left halves of the box and there are $N/2$ particles to the right of the partition and $N/2$ particles to the left. Suddenly, the partition breaks allowing the particles on the right and left to freely mix. If the particles are indistinguishable, there is no change in the entropy of the system when the partition breaks. However, if the particles are distinguishable (for instance, if the right and left halves contain different isotopes of the same inert gas) then there is an increase in entropy, called the entropy of mixing, equal to $\Delta S = kN \ln 2$. Prove these statements.

In the nineteenth century, before the development of quantum mechanics and the notion of indistinguishability, physicists thought that identical particles were distinguishable and that there should therefore be an increase in entropy when identical gases mix. Experimentally, no such increase was observed and this puzzling state of affairs was called Gibbs's paradox.

SECTION 15.7 (The Quantum Ideal Gas — A Many-Particle System)

15.37 • Use the technique of separation of variables to verify that the wave functions (15.22) with energies (15.23) are solutions of the Schrödinger Eq. (15.20).

15.38 •• As in Fig. 15.10(b), the ground state of a gas of fermions consists of one-eighth of a sphere of states in k space, with all states filled up to a maximum energy, called the *Fermi energy*. (a) For a gas of N spin- $\frac{1}{2}$ noninteracting fermions in a container of volume $V = a^3$, derive an expression for the Fermi energy. (b) Compute the approximate value of the Fermi energy in eV for the case of an electron gas in a metal. Assume one conduction electron per atom, and a lattice constant (distance between nearest-neighbor atoms) of 0.3 nm.

15.39 •• Show that for a degenerate gas of fermions at $T = 0$, the average energy per particle and the Fermi energy E_F (the energy of the highest occupied states) are related by $\langle E \rangle = (3/5)E_F$. [Hint: Avoid writing any messy constants in your calculation, by noting that we can write

$$\langle E \rangle = \frac{\int_0^{E_F} p(E)E dE}{\int_0^{E_F} p(E) dE}$$

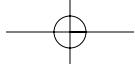
where $p(E) = C\sqrt{E}$ and C is a constant that you don't need to know because it cancels.]

15.40 ••• Consider a quantum ideal gas of N spinless particles in a container of volume $V = a^3$. If the temperature is sufficiently high and the density is sufficiently low, the gas is in the classical nondegenerate regime, in which the probability that any particular point in k space is occupied is much less than one, corresponding to Fig. 15.12(a). In this high-temperature, low-density limit, the properties of the gas are the same regardless of whether the particles of the gas are fermions or bosons. (a) Show that this classical regime occurs when the number density (number per volume) n obeys $n \ll (3kTm)^{3/2}/6\pi^2\hbar^3$, where m is the mass of a gas molecule. [Hint: From the equipartition theorem, the average kinetic energy of a single particle is $\frac{3}{2}k_B T$, and so a typical value of $k = |\mathbf{k}|$ for a particle is given roughly by $\hbar^2 k^2 / 2m \approx \frac{3}{2}k_B T$. Also, recall that the density of points in k space is 1 point per $(\pi/a)^3$.] (b) For a gas of helium atoms at atmospheric pressure, below what temperature (roughly) does the gas become degenerate? (c) For a conduction electron gas in a metal, how high a temperature would be necessary for the electrons to be in the classical regime? Is such a temperature possible? Assume 1 conduction electron per atom, and a volume per atom of about $(0.3\text{ nm})^3$.

SECTION 15.8 (Energy and Speed Distributions in an Ideal Gas)

15.41 • (a) In an ideal monatomic gas at temperature T , what is the ratio of the probabilities that a particle has speeds $2v$ and v , that is, what is $p(2v)/p(v)$?

- (b)** For a gas of helium atoms at $T = 300\text{ K}$, what is this ratio when v is the rms speed $\sqrt{3kT/m}$?
- 15.42 ••** Protons on the surface of the sun can be considered to be a gas of noninteracting particles. **(a)** Calculate the escape speed from the surface of the sun. ($M_{\text{sun}} = 1.99 \times 10^{30}\text{ kg}$, $R_{\text{sun}} = 6.96 \times 10^8\text{ m}$). **(b)** Assume that the protons at the sun's surface are in thermal equilibrium at $T = 6000\text{ K}$. Argue that none of the protons on the surface of the sun has a speed greater than the escape speed. (See, however, Problem 15.67.)
- 15.43 ••** Eq. (15.29) is the distribution of single-particle energies in a three-dimensional ideal gas. Derive the corresponding expression for the case of a two-dimensional ideal gas. [Hint: Instead of considering a shell of states in 3-D k space, consider a ring of states in 2-D k space.] (A physical example of a 2-D gas is the electron gas that forms at the interface between doped GaAs layers in some semiconductor devices.)
- 15.44 ••** **(a)** At what value of the molecular speed v does the Maxwell–Boltzmann distribution of speeds have it maximum? **(b)** Show that this *most probable speed* corresponds to a kinetic energy kT .
- 15.45 ••** In an ideal gas, at a given instant in time, roughly what fraction of the molecules have average kinetic energies $\text{KE} < 0.1 kT$? [Hint: If $E \ll kT$, then $e^{-E/kT} \approx 1$.]
- 15.46 ••** The standard deviation of a random variable x is defined as $\Delta x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$. Note that this is a measure of the “spread” of values of x about the mean value. **(a)** Show that the standard deviation can also be written as $\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$. **(b)** What is the standard deviation Δv of the distribution of speeds of an atom in an ideal monatomic gas at temperature T ? **(c)** What is the ratio $\Delta v/v_{\text{rms}}$?
- 15.47 ••** Prove that for the particles of an ideal monatomic gas, $\langle E \rangle = \int_0^\infty Ep(E) dE = \frac{3}{2}kT$, where $p(E)$ is the energy distribution given by (15.31).
- 15.48 ••** Prove that for an ideal gas, $\langle v \rangle = \int_0^\infty vp(v) dv = \sqrt{8kT/(\pi m)}$, where $p(v)$ is the Maxwell speed distribution given by (15.32).
- 15.49 ••** Derive expressions for $\langle v \rangle^2$ and $\langle v^2 \rangle$ in an ideal gas, where v is the speed of the molecules. (The equipartition theorem and probability distribution $p(v)$ given by (15.32) will be useful here.) Which expression is larger, and why?
- SECTION 15.9 (Heat Capacities)**
- 15.50 • (a)** How much energy is required to raise the temperature of a mole of an ideal monatomic gas by 1°C at constant volume? Give your answer in both J/mol and eV/atom . **(b)** How much energy is required to raise the temperature of 1 kg of aluminum by 1°C at room temperature (where the Dulong–Petit law applies)?
- 15.51 ••** Prove the two identities in (15.34):
- $$\frac{1}{(1-x)} = 1 + x + x^2 + x^3 + \dots$$
- and
- $$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \dots$$
- 15.52 ••** Show that the average energy of a system can be written as $\langle E \rangle = -d(\ln Z)/d\beta$, where $\beta = 1/kT$ and Z is the partition function $Z = \sum_i g_i e^{-\beta E_i}$.
- 15.53 ••** Show that the heat capacity of a system can be written as $C = Td^2(kT \ln Z)/dT^2$, where Z is the partition function $Z = \sum_i g_i e^{-\beta E_i}$. [Hint: Use the results of Problem 15.52.]
- 15.54 ••** Consider a three-level quantum system, with non-degenerate energy levels 0, $+\varepsilon$, and $+2\varepsilon$. (A physical example of such a system is a spin $S = 1$ magnetic moment in a magnetic field.) **(a)** Write down the partition function Z for this system. **(b)** Write down the expression for the thermal average energy $\langle E \rangle$ of this system. **(c)** Show that in the low-temperature limit $kT \ll \varepsilon$, the energy and specific heat of this system are identical to those of a two-level system. **(d)** Show that in the high-temperature limit $kT \gg \varepsilon$, the heat capacity of this system, like that of a two-level system, approaches zero.
- 15.55 ••** The argument leading to the Dulong–Petit law treats the lattice vibrations classically, that is, the spectrum of energies is assumed to be continuous. This would be expected to break down once the thermal energy ($\sim k_B T$) is comparable to, or less than, the quantized level spacing $\Delta E = \hbar\omega_c$ of the lattice vibrations. Given that the latter is of order 0.03 eV, estimate the temperature below which the Dulong–Petit law should fail. Compare your answer with Fig. 15.17.
- 15.56 ••** The Dulong–Petit law predicts that the molar specific heats (that is, the heat capacity *per mole*) of elemental solids should all have the same value, $C = 3R$. On the other hand, the ordinary specific heats c , defined as the energy per degree *per mass*, can be quite different. **(a)** Prove that the specific heats, c , of elemental solids obeying the Dulong–Petit law are inversely proportional to the atomic masses. **(b)** Given the following data, check the accuracy of this prediction.
- | Solid: | Aluminum | Silver | Gold |
|---|----------|--------|--------|
| $c[\text{cal}/(\text{g} \cdot \text{K})]$: | 0.214 | 0.0558 | 0.0312 |
- 15.57 •• (a)** Assuming their molar specific heats are given by the Dulong–Petit law, predict the specific heats, in $\text{cal}/(\text{g} \cdot \text{K})$, of Al, Ag, and Au. **(b)** Find the percent discrepancies between your predictions and the measured values given in Problem 15.56.



532 Chapter 15 • Statistical Mechanics

15.58 •• Compare the measured specific heat of ice [$c_{\text{ice}} = 0.49 \text{ cal}/(\text{g} \cdot \text{K})$] at temperatures somewhat below freezing] with that expected from the Dulong–Petit law. From this comparison, can you estimate how many internal degrees of freedom of the ice molecule are contributing to the heat capacity?

15.59 •• Consider a dilute gas of N noninteracting atoms, in which the atoms have a nondegenerate ground state with energy $E_0 = 0$ and a first excited level that has an energy $E_1 = \varepsilon$ and degeneracy g . Assume that the second excited state has an energy very high compared with the thermal energy; that is, assume $kT \ll E_2$, so that the level E_2 and all higher levels can be ignored. **(a)** At temperature T , what is the ratio of the number of atoms in the first excited level (degeneracy g) to the number in the ground state? **(b)** What is the average energy of an atom in this gas? **(c)** What is the total energy of the gas? **(d)** What is the specific heat of the gas?

15.60 ••• Consider a simple solid, such as aluminum, consisting of N identical atoms. Assume that the Einstein approximation correctly describes the vibrational motion of the atom. **(a)** Assume that the vibrational level spacing has the value $\varepsilon = 0.037 \text{ eV}$ (a typical value in a solid). Consider the probability that an atom of the solid is in its ground state. At what temperature is this probability equal to 0.1? **(b)** For a given N and ε , at what temperatures is the average number of atoms in the ground state less than one? That is, how high does the temperature have to be to ensure that all atoms are in an excited state? Is this temperature achievable in an ordinary sample of a solid?

15.61 •• Make an order-of-magnitude estimate of the temperature at which the internal vibrational mode of a nitrogen molecule becomes appreciably excited. Recall that the energy of the first excited state of a simple harmonic oscillator is $\hbar\omega$, where $\omega = \sqrt{k/m}$ is the classical angular frequency, k is the spring constant, and m is the mass. Recall also that the energy of the oscillator can be written as $U = \frac{1}{2}kx^2 = \frac{1}{2}m\omega^2x^2$. Finally, note that we can expect a typical chemical bond energy $U(x) \sim 3 \text{ eV}$ when x is a typical atomic distance $x \approx 0.1 \text{ nm}$. The moral of this question is that the vibrational modes of air molecules are largely “frozen out” at room temperature.

15.62 •• Make an order-of-magnitude estimate of the temperature at which the rotational modes of a nitrogen molecule become excited. Recall that the kinetic energy of a rotating system is $U = \frac{1}{2}I\omega^2 = \frac{1}{2}L^2/I$, where $I = \sum m_i r_i^2$ is the moment of inertia and $L = I\omega$ is the angular momentum. Recall that for a quantum system, the first excited state has $L^2 \approx \hbar^2$. Based on your answer, would you expect the rotational modes of air molecules to contribute to the heat capacity of air at room temperature?



COMPUTER PROBLEMS

15.63 •• (Section 15.4) For very large numbers, the factorial of N can be estimated by *Stirling's approximation* $N! \approx (\sqrt{2\pi N})N^Ne^{-N}$, which can also be written as $\ln(N!) \approx \frac{1}{2}\ln(2\pi N) + N\ln N - N$. Often, to simplify calculations, a further approximation is made and the formula is written as $\ln(N!) \approx N\ln N - N$. **(a)** Plot $\ln(N!)$ versus N for $N = 1, 2, \dots, 10$, and on this same plot, graph the long version of Stirling's approximation. **(b)** Make a graph showing how the long and short versions of Stirling's approximation compare over the range of 10 to 100. Consider carefully what plot will show this comparison to best effect.

15.64 •• (Section 15.8) Reproduce Figure 15.14, the distribution of single-particle energies in an ideal gas, for the case of helium gas at room temperature.

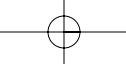
15.65 •• (Section 15.8) Reproduce Figure 15.15, showing the Maxwell speed distributions, for the cases of N_2 and He gases at room temperature.

15.66 •• (Section 15.8) The temperature of the uppermost reaches of the earth's atmosphere, called the thermosphere, is approximately 600 K during the day (the high temperature is due to efficient absorption of solar radiation). **(a)** At 600 K, what fraction of hydrogen molecules have thermal speeds greater than the earth's escape velocity? **(b)** What fraction of the hydrogen molecules have speeds greater than the escape velocity at room temperature, about 300 K? [Hint: You need the computer to do the required integration.]

15.67 •• (Section 15.8) The outer atmosphere of the sun, called the corona, consists mostly of protons and extends out to distances of two to three times the radius of the sun. The temperature of the corona is extraordinarily high, approximately 1 million degrees Kelvin (the cause of this high temperature is still not well understood). **(a)** Compute the approximate escape speed for particles in the outer corona. ($M_{\text{sun}} = 1.99 \times 10^{30} \text{ kg}$, $R_{\text{sun}} = 6.96 \times 10^8 \text{ m}$) **(b)** Approximately what fraction of the protons in the outer corona have a thermal speed greater than the escape speed? [Protons escaping from the corona cause the “solar wind.”]

15.68 •• (a) (Section 15.9) Plot the Einstein specific heat of a solid as a function of temperature and thus reproduce Figure 15.17. Assume that the vibrational level spacing has the value $\varepsilon = 0.037 \text{ eV}$ (a typical value in a solid). **(b)** Using the value $\varepsilon = 0.037 \text{ eV}$, at what temperature is the heat capacity of the solid given by 0.9 times the Dulong–Petit value?

15.69 •• (Section 15.9) Reproduce Figure 15.18, showing the average energy and heat capacity, for the case of a three-level system with nondegenerate energy levels 0, $+\varepsilon$, and $+2\varepsilon$.



SUBATOMIC PHYSICS

PART IV



- Chapter 16 The Structure of Atomic Nuclei
- Chapter 17 Radioactivity and Nuclear Reactions
- Chapter 18 Elementary Particles

In **Parts I and II** we developed the two theories, relativity and quantum mechanics, that revolutionized twentieth-century physics. In Part III we described some of the many applications of these theories to systems containing several atoms — molecules, solids, and some of the systems treated by statistical mechanics. Now, in Part IV we discuss subatomic systems, namely, atomic nuclei and subnuclear particles. As far as practical, we have made the chapters of Parts III and IV independent of one another so that you can select just those topics that interest you. In particular, you do not have to have read Part III at all before reading Part IV.

Part IV contains three chapters. Chapters 16 and 17 describe the properties of atomic nuclei, while Chapter 18 treats the subnuclear, “elementary” particles. The division between the two chapters on nuclear physics is that Chapter 16 is concerned with the properties of an isolated, static nucleus, while Chapter 17 treats nuclear reactions, in which nuclei change their energy and often their identity as well. These two chapters are not completely independent of one another, but if your main interest is in nuclear reactions (Chapter 17), you need to read only Sections 16.1 to 16.3 before skipping to Chapter 17. If your main interest is in particle physics (Chapter 18), you could read just Sections 16.1 to 16.3 and 17.1 to 17.4 before skipping to Chapter 18.

Chapter 16

The Structure of Atomic Nuclei

- 16.1** Introduction
- 16.2** Nuclear Properties
- 16.3** The Nuclear Force
- 16.4** Electrons Versus Neutrons as Nuclear Constituents
- 16.5** The IPA Potential Energy for Nucleons
- 16.6** The Pauli Principle and the Symmetry Effect
- 16.7** The Semiempirical Binding-Energy Formula
- 16.8** The Shell Model*
- 16.9** Mass Spectrometers*

Problems for Chapter 16

*These sections can be omitted without loss of continuity.

16.1 Introduction

In discussing the properties of atoms, we needed to know surprisingly little about the atomic nucleus. In fact, for most purposes we needed only three facts: The nucleus carries a positive charge Ze ; the nucleus is so much heavier than the electrons that we can usually treat it as fixed; and the nucleus is so small that — for the purpose of atomic physics — we can usually treat it as a point particle.

In this and the next chapter we turn our attention to the internal structure of atomic nuclei. We will find that in some ways the motion of the protons and neutrons inside a nucleus is analogous to the motion of the electrons in an atom. Many of the ideas developed for atomic physics (the Schrödinger equation, the Pauli principle, spin, etc.) are equally useful in nuclear physics.

Nevertheless, there are significant differences between atomic and nuclear physics, of which perhaps the most important is the difference between the forces involved. The force that holds electrons in their atomic orbits is the familiar electrostatic force, which was well understood long before the advent of atomic physics. In nuclear physics, the situation was quite different, at least historically. As we will argue in Section 16.3, the force that holds protons and neutrons together in a nucleus is not a force that was known in classical physics, and even today we do not have a complete theoretical understanding of nuclear forces. Therefore, the history of nuclear physics during the last 80 years is the story of a simultaneous effort to elucidate the nuclear force and to explain the properties of nuclei. In this sense, the development of nuclear

physics has been more difficult than that of atomic physics. Nevertheless, we will see that nuclear physicists have achieved a remarkable understanding of nuclear properties. We will find that this understanding is both interesting for its own sake and of great practical importance.

16.2 Nuclear Properties

We begin our study of nuclear physics with a survey of several nuclear properties, some of which we met briefly in Chapter 3. In later sections we will see that many of these properties can be understood with the help of the same quantum principles developed in Chapters 6 through 10.

Constituents

We have already stated that nuclei are made up of protons and neutrons. As we will describe in Chapter 18, these two kinds of particles are themselves made up of quarks, but for the present we can treat protons and neutrons as “elementary” particles. We denote the number of protons by Z , and the number of neutrons by N . Since each proton has charge $+e$, while neutrons are uncharged, the total charge of the nucleus is $+Ze$.

The mass of a nucleus is

$$m_{\text{nuc}} = Zm_p + Nm_n - \frac{B}{c^2} \quad (16.1)$$

where B is the total binding energy, the energy required to pull the nucleus apart into its $Z + N$ separate constituents. As discussed in Chapter 3, the binding-energy term, $-B/c^2$, is usually less than 1% of m_{nuc} . Furthermore, the difference between the masses m_p and m_n is only 1 part in 1000. Hence, we can often neglect the last term in (16.1) and set $m_n \approx m_p$ to give the approximation

$$m_{\text{nuc}} \approx (Z + N)m_p = Am_p$$

where, as before, we have defined the **mass number**, or **nucleon number**,

$$A = Z + N$$

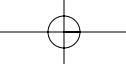
which is the total number of nucleons (protons *and* neutrons) in the nucleus. It is sometimes convenient to write some (and occasionally all) of the numbers A , Z , and N explicitly with the chemical symbol. The agreed placement around the chemical symbol X is now*

$$_Z^A X_N$$

Some examples are

$${}_1^1 H_0, {}_1^2 H_1, {}_2^4 He_2, {}_{12}^{12} C_6, {}_{29}^{63} Cu_{34}, {}_{92}^{238} U_{146} \quad (16.2)$$

*This arrangement has changed from time to time, and many older books use different conventions. On those relatively rare occasions when we need to use the right subscript to indicate the neutron number, be careful not to confuse it with the right subscript used to indicate the number of atoms in a molecule.



4 Chapter 16 • The Structure of Atomic Nuclei

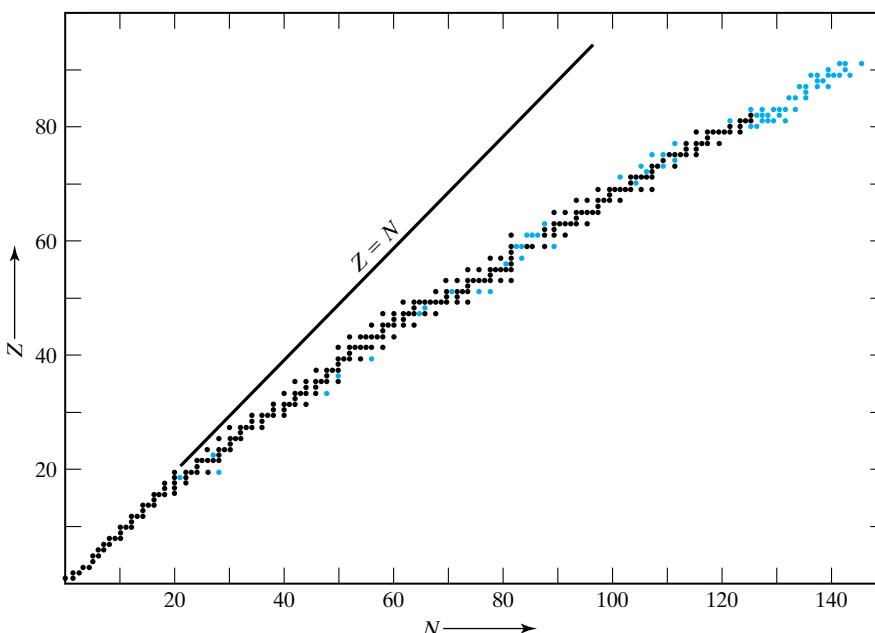


FIGURE 16.1

The values of N and Z for all naturally occurring nuclei. Stable nuclei are shown by black dots; unstable, by colored dots.

Fortunately, it is usually not necessary to give all of the three numbers on these symbols, and we will generally include only as many as seem necessary for the discussion at hand. (This may be a good time to remind you that the first three nuclei in (16.2) have their own special names: ${}^1\text{H}$ is, of course, the proton, denoted p ; the nucleus of heavy hydrogen, ${}^2\text{H}$, is called the deuteron, and is sometimes denoted D ; and the nucleus ${}^4\text{He}$ is called the alpha particle, denoted α .)

Two different nuclei, such as ${}^{16}_8\text{O}_8$ and ${}^{17}_8\text{O}_9$, that have the same number of protons, are called **isotopes**. As discussed in Chapter 3, the corresponding atoms are chemically indistinguishable, or very nearly so. Two nuclei, such as ${}^{14}_6\text{C}_8$ and ${}^{14}_7\text{N}_7$, with the same mass number are called **isobars**; and two nuclei, such as ${}^{13}_6\text{C}_7$ and ${}^{14}_7\text{N}_7$, with the same number of neutrons are called **isotones**.

The naturally occurring elements have atomic numbers in the range $1 \leq Z \leq 92$ and neutron numbers with $0 \leq N \leq 146$. The values of Z and N for all naturally occurring nuclei are plotted in Fig. 16.1, from which one sees that for light nuclei ($A \lesssim 40$) N and Z are nearly equal; while for heavier nuclei, N is a little larger than Z , with $N \approx 1.5Z$ in the heaviest nuclei. The tendency to have $N \approx Z$ is called the *symmetry effect*. We will show in Section 16.6 that the symmetry effect results from the Pauli principle and the so-called charge independence of nuclear forces and that the tendency for N to be somewhat bigger than Z is due to the electrostatic repulsion of the protons in nuclei with large Z .

Another important tendency among nuclei is the observed preference for Z and N to be even, as illustrated in Table 16.1. The stable nuclei in which both Z and N are even outnumber comfortably those in which either Z or N is odd, and there are only four stable nuclei in which *both* Z and N are odd. This tendency is the result of the so-called “pairing effect,” which is described in Section 16.8.

Sizes and Shapes of Nuclei

We described in Chapter 3 how Rutherford probed the atomic nucleus using energetic alpha particles. (The main reason that he used alpha particles was simply that they were readily available, since they are ejected with various different

TABLE 16.1

Numbers of stable nuclei.

Z	N	Number of stable nuclei
Even	Even	148
Even	Odd	51
Odd	Even	49
Odd	Odd	4

energies by many naturally occurring radioactive substances.) Rutherford's experiments established that nuclei are much smaller than atoms, with radii of just a few femtometers, or fermis ($1 \text{ fm} = 10^{-15} \text{ m}$). More recent experiments have used artificially accelerated beams of electrons, protons, and alpha particles to probe the size and shape of atomic nuclei. The results of these experiments are summarized in the next few paragraphs.

The majority of nuclei are spherical. A number are slightly nonspherical, some being prolate (elongated, like an American football), and some oblate (flattened, like most pumpkins). In nonspherical nuclei the longest diameter is typically a few percent greater than the shortest diameter, although in a few cases this difference can be as large as 30%. It is usually a reasonable approximation to treat nonspherical nuclei as spherical, and, for the most part, we will consider nuclei to be spherical.

The radii of nuclei increase steadily from about 2 fm for light nuclei (helium, for example) to about 7 fm for heavy nuclei (uranium, for example). It is found that the radii of all nuclei are given approximately by the simple formula

$$R = R_0 A^{1/3} \quad (16.3)$$

where A is the number of nucleons and R_0 is a constant,

$$R_0 = 1.07 \text{ fm} \quad (16.4)$$

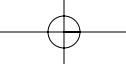
The significance of (16.3) is clearer if we consider the *volume* of the nucleus,

$$\begin{aligned} V &= \frac{4}{3}\pi R^3 = \frac{4}{3}\pi R_0^3 A \\ &= V_0 A \end{aligned} \quad (16.5)$$

where V_0 is the constant $\frac{4}{3}\pi R_0^3$. Evidently, (16.3) implies that the volume of a nucleus is proportional to A , the number of constituent nucleons. This means that the volume *per nucleon* V_0 is approximately the same in all nuclei. In other words, the density of nucleons inside the nucleus is the same in all nuclei.

The volume of a single isolated nucleon turns out to be about $V_0/2$. Thus we can say that nucleons in a nucleus take up half of the total volume. This means that the nucleons are quite closely packed inside a nucleus, quite like the molecules in a liquid — and entirely unlike the electrons in an atom.*

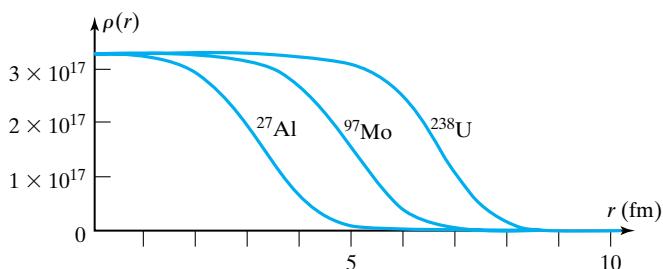
*Of course, because of their wave functions, the electrons' probability distribution is spread out through the atom. This does not change the fact that the electron itself is very small compared with the atom. In fact, current theory holds that the electron is a point particle — unlike nucleons, which have a measurable size.



6 Chapter 16 • The Structure of Atomic Nuclei

FIGURE 16.2

The density, in kg/m^3 , as a function of r in the nuclei of aluminum, molybdenum, and uranium. These graphs can be accurately fitted by an analytic expression called the Fermi function (see Problems 16.15, 16.63, and 16.64).



The radius of a nucleus is not an exactly defined quantity. This is evident in Fig. 16.2, which shows the distribution of mass in three representative nuclei. In no case is there a unique radius at which the density falls exactly to zero. The usual definition of R and the definition assumed in (16.3) and (16.4) is the radius at which the density $\rho(r)$ falls to half its maximum value $\rho(0)$. Notice that, as expected, the central density is the same in all three pictures. Notice also that all three graphs fall to zero at about the same rate; that is, the thickness of the surface region is about the same in all three nuclei.

The actual density inside a nucleus is seen from Fig. 16.2 to be about $3 \times 10^{17} \text{ kg}/\text{m}^3$, some 14 orders of magnitude greater than the density of ordinary liquids or solids. (Remember that water has density $10^3 \text{ kg}/\text{m}^3$.) The enormous density inside nuclei is easy to understand if we recall that almost all the mass of an atom is in the nucleus, whose volume is some 10^{14} times smaller than the atom (Problem 16.9). Therefore, the nuclear density is 10^{14} times bigger than the atomic density, which is itself about the same as that of ordinary liquids or solids.

Nuclear Energy Levels

Like atoms, nuclei are found to have quantized energy levels. However, the spacing of the levels in nuclei is many times larger than that in atoms. In heavy nuclei the level spacing is usually several tens of keV, while in light nuclei it can range up to 10 MeV and more. Because their excitation energies are so much larger than the energies involved in chemical reactions, nuclei remain locked in their ground states during normal chemical processes. This is why it is often possible in atomic physics to treat the nucleus as a structureless body with no internal motion.

If a nucleus is raised to one of its excited states (when struck by an energetic proton, for instance), it can return to the ground state by emitting a photon. This is analogous to the emission of photons in atomic transitions, the only important difference being that the photons emitted by nuclei are usually far more energetic. Photons emitted in nuclear transitions (with energies ranging from keVs to MeVs) are called γ -rays. Measurement of their energy tells us the separation of the two nuclear energy levels involved.

Angular Momenta of Nuclei

Like the electron, both the neutron and proton have spin half. Thus the total angular momentum \mathbf{J} of a nucleus is the sum of the orbital angular momenta of all of its nucleons *plus* all of their spins. As with all angular momenta, the magnitude of \mathbf{J} is quantized; the possible values are labeled by a quantum number j , with

$$J = |\mathbf{J}| = \sqrt{j(j+1)\hbar} \quad (16.6)$$

16.3 The Nuclear Force

The force that holds electrons in their atomic orbits is the familiar electrostatic attraction between the positive nucleus and the negative electrons. On the other hand, it should be clear that electrostatic attraction is *not* what holds the nucleons in the nucleus. First, the neutron is uncharged and feels no electrostatic force at all. Second, the protons are all positive and the effect of the electrostatic force is therefore to push the protons apart.

Since the gravitational force acts on all matter, charged or not, and is always attractive, one might wonder if gravity could be the force that holds the nucleus together. However, gravity is many orders of magnitude too weak, as the following example illustrates.

Example 16.1

Compare the gravitational attraction between two protons with their electrostatic repulsion.

The gravitational attraction is

$$F_{\text{grav}} = \frac{Gm_p^2}{r^2}$$

while the repulsive electrostatic force is

$$F_{\text{elec}} = \frac{ke^2}{r^2}$$

The ratio of these forces is independent of r and is equal to

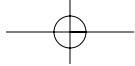
$$\frac{F_{\text{grav}}}{F_{\text{elec}}} = \frac{Gm_p^2}{ke^2} = \frac{(6.7 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2) \times (1.7 \times 10^{-27} \text{ kg})^2}{(9.0 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2) \times (1.6 \times 10^{-19} \text{ C})^2} \approx 10^{-36}$$

Clearly, the gravitational force is far too weak to overcome the electrostatic repulsion of the protons and hold the nucleus together.

Since neither the electrostatic force nor gravity holds the nucleus together, there must be some other force between nucleons, and we call it the **nuclear force**. The nuclear force must be stronger than the electrostatic force, in the sense that it must more than balance the electrostatic repulsion of the protons. For this reason, the nuclear force is often called the **strong force**.

A useful way to view the strength of the nuclear force is in terms of the energy of a nucleon in the nucleus. As a crude model, we can imagine any one nucleon moving inside a cubical box whose side a is of order 6 fm (the approximate diameter of a fairly light nucleus such as aluminum, ^{27}Al). We saw in Chapter 8, Equation (8.32), that the minimum kinetic energy for such a particle is

$$\begin{aligned} K_{\min} &\approx \frac{3\hbar^2\pi^2}{2ma^2} = \frac{3\pi^2(\hbar c)^2}{2mc^2a^2} \\ &\approx \frac{3\pi^2(200 \text{ MeV} \cdot \text{fm})^2}{2 \times (940 \text{ MeV}) \times (6 \text{ fm})^2} \approx 20 \text{ MeV} \end{aligned} \tag{16.7}$$



8 Chapter 16 • The Structure of Atomic Nuclei

That is, the minimum kinetic energy of any one nucleon in a nucleus is of order 20 MeV. In order that the nucleus hold together, its total energy must be negative. Therefore, the nuclear force on any nucleon due to the other $A - 1$ nucleons must produce a potential-energy well of depth at least several tens of MeV.

While the nuclear potential well must be very deep, it must also be of very short range; that is, the force between nucleons must fall rapidly to zero as their separation increases. Perhaps the simplest evidence for this comes from experiments in which a beam of protons is scattered by nuclei. Those protons that pass by a nucleus at large distance are influenced only by the Coulomb repulsion between the proton's charge and the nuclear charge. Even protons passing within 2 or 3 fm of the nuclear surface scatter in the pattern predicted on the basis of a simple Coulomb repulsion. Only when the protons approach within 1 or 2 fm of the nuclear surface does the pattern of scattering change, indicating the presence of a strong attractive nuclear force. We conclude that the nuclear force between nucleons is negligibly small when they are more than 2 or 3 fm apart.

At separations of 1 or 2 fm, the force between nucleons is strongly attractive, as we have just argued. However, at separations less than about 1 fm the nuclear force becomes strongly repulsive. The simplest argument for this last claim is that, as we saw in Section 16.2, the density inside a nucleus is approximately constant and is the same for all nuclei. If the nuclear force were purely attractive, we would expect nucleons at the center of a large nucleus to be forced inward more tightly than in a small nucleus. The easiest way to explain the observed constancy of nuclear densities is to assume that at very short distances the nucleons are held apart by a repulsive force, and this proves to be the case.*

Today, much of the best evidence for the details of the nuclear force comes from nucleon-nucleon scattering experiments, in which a beam of neutrons or protons is scattered by the protons in a hydrogen target, much as Rutherford's alpha particles were scattered by the nuclei in a metal foil. The main conclusions of these experiments are summarized in Fig. 16.3, which shows the potential energy of the nuclear force between two nucleons, plotted as a function of their separation r . You can see that the main features of this graph agree with our qualitative conclusions: The nuclear potential energy approaches zero rapidly beyond about 2 fm. Inside 2 fm the force is attractive and the potential energy becomes negative, dropping to a minimum of about -100 MeV at about 1 fm. Inside this minimum, the force is repulsive and the potential energy rapidly becomes positive and large.[†]

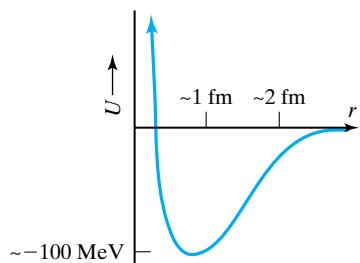


FIGURE 16.3

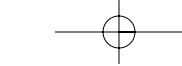
The nuclear potential energy of two nucleons as a function of their separation r .

Charge Independence of Nuclear Forces

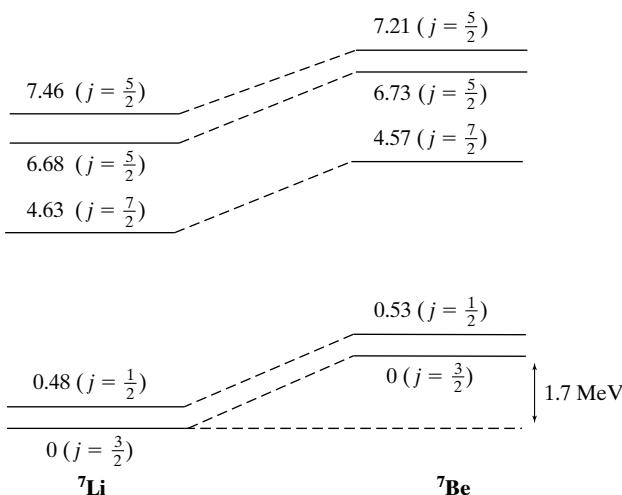
Two final properties of the nuclear force should be mentioned. First, it is found that for a given quantum state, the nuclear force between two nucleons is the same whether they are both protons, or both neutrons, or one proton and one neutron. This property is called the **charge independence of nuclear forces** — the nuclear force on a nucleon is independent of whether the nucleon is charged (a proton) or uncharged (a neutron). It is charge independence that is

*In fact, the constancy of the density in nuclei involves several other effects in addition to the short-range repulsion. Nevertheless, the conclusion stated is correct.

[†] Strictly speaking, the potential energy sketched in Fig. 16.3 is that for two nucleons with zero angular momentum. The force between nucleons depends somewhat on their angular momentum, but for our present purposes, the potential energy shown is the most important.



Section 16.4 • Electrons Versus Neutrons as Nuclear Constituents 9

**FIGURE 16.4**

Energy levels of the isobars ${}^7\text{Li}_4$ and ${}^7\text{Be}_3$. The numbers labeling each level are its excitation energy in MeV and its angular-momentum quantum number j . Corresponding levels are connected by a dashed line; they have angular momenta that are exactly equal and excitation energies that are very nearly so. This close agreement is evidence for the charge independence of nuclear forces.

responsible for the observed tendency for nuclei to have equal numbers of neutrons and protons as we describe in Section 16.6.

Much detailed evidence for the charge independence of nuclear forces comes from studies of the interactions between two nucleons (p–p, p–n, n–n) in collision experiments. Further evidence comes from the study of energy levels in nuclear *isobars*, that is, two or more nuclei that have the same total number of nucleons and differ only in their numbers of protons and neutrons. Consider, for example, the pair of nuclei ${}^7\text{Li}_4$ and ${}^7\text{Be}_3$. The only difference between these is that one neutron in ${}^7\text{Li}_4$ has been replaced by a proton in ${}^7\text{Be}_3$. Figure 16.4 shows the lowest five energy levels of each of these two nuclei. Because ${}^7\text{Be}$ has one extra charge, all of its levels are raised by 1.7 MeV due to the additional Coulomb repulsion. Apart from this obvious difference, the levels of the two nuclei are almost identical. The close correspondence of the energy levels (both in excitation energy and angular momentum) is strong evidence that the nuclear force acts equally on neutrons and protons.

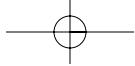
Electrons and the Nuclear Force

Although the nuclear force is so crucially important for neutrons and protons, it does not act on electrons at all — in just the same way that the electrostatic force does not act on particles that are electrically neutral. The principal evidence for this claim comes from experiments in which high-energy electrons collide with nuclei; even when the electrons penetrate deep into the nucleus, there is no evidence that they are affected by the strong nuclear force. This makes the electron a very useful probe of nuclei, since its interaction with the nucleons is just the well-understood electromagnetic interaction.

16.4 Electrons Versus Neutrons as Nuclear Constituents

With our general knowledge of nuclear properties and forces, we are now ready to examine several questions in more detail. The first question that we address concerns the constituents of nuclei.

Although we now know that nuclei consist of protons and neutrons, this was not always obvious. That nuclei contain protons was established in 1919



10 Chapter 16 • The Structure of Atomic Nuclei

when Rutherford observed protons ejected from collisions between alpha particles and nitrogen nuclei. The neutron was not discovered for another 12 years, and at first it was natural to suppose that nuclei were made of protons and *electrons*. Recall that the two basic facts about nuclei are that they have charge Ze and mass Am_p (approximately). Today we explain these by saying that the nucleus has Z protons, to give it charge Ze , and $A - Z$ neutrons (with $m_n \approx m_p$), to give it total mass Am_p . However, the same two facts could be equally well explained by supposing that the nucleus has A protons and $A - Z$ electrons. Because m_e is so small, the mass would still be about Am_p ; and because the electron has charge $-e$, the $A - Z$ electrons would reduce the total charge to Ze , as observed.

The picture of the nucleus as made of protons and electrons had two very appealing features. First, it was economical: If correct, it would have meant that all matter was made of just two kinds of particles, protons and electrons. Second, since electrons and protons have opposite charges, it would have been conceivable that the nucleus was held together by electrostatic attraction.

The question that we address here is this: How do we know that the picture of nuclei as made of protons and neutrons is correct and the proton-electron picture incorrect? We present two arguments.

The Energy of Nuclear Constituents

We have seen that the kinetic energy of a proton or neutron inside a nucleus is at least several MeV. This requires that the nuclear force be strong enough to confine nucleons with these energies, and we have seen that the observed nuclear force is sufficiently strong. Suppose, however, that the nucleus contained electrons. It is a simple matter to estimate the minimum kinetic energy of an electron in a nucleus, as in the following example.

Example 16.2

Assuming that an electron could be confined inside a nucleus of radius about 3 fm (for example, ^{27}Al), estimate the electron's minimum kinetic energy.

To estimate the kinetic energy of nucleons in a nucleus, we treated the nucleus as a rigid cubical box. We can do the same thing for an electron, but we need to be careful because the electron mass is only $0.5 \text{ MeV}/c^2$, so we must anticipate that its motion may be relativistic. Therefore, we will start from the general relation

$$E = \sqrt{(pc)^2 + (mc^2)^2} \quad (16.8)$$

and find the minimum possible value of p^2 .

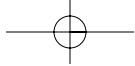
We saw in Section 8.3 that the requirement that the wave function vanish at the walls of the box led to a minimum for each component of momentum:

$$p_x = \hbar k_x = \frac{h}{2a} \quad (16.9)$$

where a is the length of the cube's edge, and the same minimum applies to p_y and p_z as well.* Therefore, the minimum possible value of p^2 is

$$p^2 = p_x^2 + p_y^2 + p_z^2 = \frac{3h^2}{4a^2}$$

*The minimum, $p_x = h/2a$, in (16.9) reflects that the maximum possible wavelength λ is $\lambda = 2a$.

Section 16.4 • Electrons Versus Neutrons as Nuclear Constituents **11**

Setting $a \approx 6 \text{ fm}$ (the diameter of the nucleus), we find that

$$(pc)^2 = \frac{3(hc)^2}{4a^2} \approx \frac{3 \times (1240 \text{ MeV} \cdot \text{fm})^2}{4 \times (6 \text{ fm})^2} \approx 3.2 \times 10^4 \text{ MeV}^2$$

Substituting into (16.8), we obtain for the minimum energy

$$E = \sqrt{(pc)^2 + (mc^2)^2} \approx \sqrt{(3 \times 10^4) + (0.25)} \text{ MeV} \approx 180 \text{ MeV}$$

Note that this answer is much larger than $mc^2 = 0.5 \text{ MeV}$, and hence that the electron would indeed be moving relativistically. Therefore, the kinetic energy, $K = E - mc^2$, is close to the total energy and

$$K_{\min} \approx 180 \text{ MeV} \quad (16.10)$$

This example shows that an electron confined inside a nucleus would have kinetic energy of order at least 100 MeV. However, there is no known force that could confine an electron with this much kinetic energy. In the 1920s the only known possibility was the electrostatic Coulomb force, and since we now know that the nuclear force does not act on electrons, the Coulomb force is still the only candidate. It is easy to estimate the electrostatic potential energy of an electron inside any given nucleus. For ^{27}Al , for example, one finds (Problem 16.25)

$$U \approx -9 \text{ MeV}$$

This energy is negative and would tend to bind the electron, but compared with (16.10) it is hopelessly too small to overcome the minimum kinetic energy of the electron, and we conclude that there can be no electrons bound inside a nucleus.

Magnetic Moments of Nuclei

Our second argument against nuclei made of protons and electrons concerns their magnetic moments. We saw in Chapter 9 that the electron has a magnetic moment whose magnitude is given by the Bohr magneton,

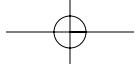
$$\mu_B = \frac{e\hbar}{2m_e} \quad (16.11)$$

where m_e is the electron's mass. (The exact magnetic moment depends on the orbital quantum number l but is always equal to μ_B times a number of order 1.) In the same way we would expect the nucleons' magnetic moments to be of order

$$\mu_N = \frac{e\hbar}{2m_N} \quad (16.12)$$

where m_N denotes the mass of the nucleon. This proves to be correct: Both the proton and neutron* have magnetic moments whose magnitudes are close to the value (16.12), which is therefore called the **nuclear magneton**.

*Since the neutron is chargeless, one might expect it to have no magnetic moment. However, we now know that although its total charge is zero, there are charges inside the neutron, and it is the motion of these charges that gives the neutron its magnetic moment.



12 Chapter 16 • The Structure of Atomic Nuclei

The important difference between μ_B in (16.11) and μ_N in (16.12) is that, since m_e is about 2000 times *smaller* than m_N , the Bohr magneton μ_B is 2000 times *greater* than μ_N . Thus if there were any electrons in the nucleus, their large magnetic moment would dominate, and the total magnetic moment would be of order μ_B . In fact, however, the measured magnetic moments of all nuclei are far smaller than this and are all of order μ_N — in keeping with the proton–neutron theory.

For these and other reasons (Problems 16.23 and 16.24) it is clear that there are no electrons inside the atomic nucleus, which consists, instead, of protons and neutrons.

16.5 The IPA Potential Energy for Nucleons

In studying atomic structure we relied heavily on the independent-particle approximation, or IPA.* In this approximation each of the Z electrons was assumed to move independently in the field produced by the average distribution of the other $Z - 1$ electrons. Working within this approximation, we found the possible states of each separate electron and, hence, those of the atom as a whole. In this way, we were led to the shell structure of atoms and to a remarkably successful account of the periodic properties of the elements.

The question naturally arises whether the corresponding approximation would be useful in nuclear physics. In the early days of nuclear physics it was generally believed that the answer would be no. The main reasons were that the nuclear force is so strong and of such short range. Figure 16.5(a) is a schematic representation of the potential energy of a nucleon moving along the x axis and interacting with just one other nucleon. (This is just the graph of Fig. 16.3 folded over, because x — unlike r — can be positive or negative, corresponding to the nucleon being to the right or left of its fixed companion.) Figure 16.5(b) shows the potential energy of one nucleon in a one-dimensional nucleus containing four other nucleons. The dashed curve shows a smoothed, average potential energy, and at first sight it would seem an extremely bad approximation to replace the rapidly fluctuating solid curve by the smooth dashed curve.

In fact, however, the effect of the rapidly fluctuating nuclear potential energy is approximated surprisingly well by its smoothed IPA average. There are several reasons for this, some of which depend on subtle cancellations between the attractive and repulsive parts of the nuclear force. However, one simple reason is just that the system in question is a quantum system. Figure 16.5(b) represents the potential energy of a nucleon in the field of four classical nucleons, each at a definite position in the nucleus. But quantum mechanics tells us that the four nucleons are themselves spread out in accordance

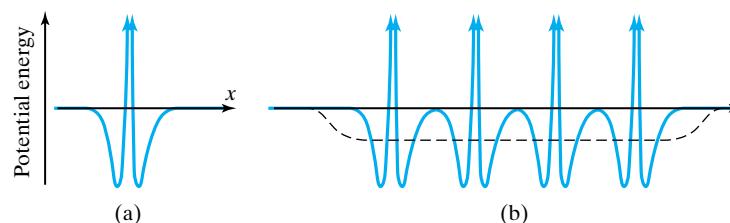
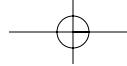


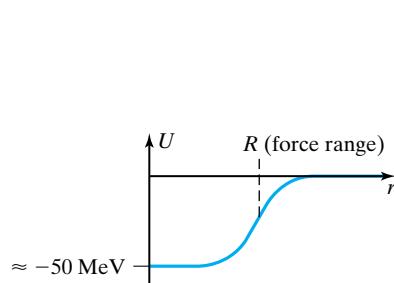
FIGURE 16.5

(a) The nuclear potential energy of one nucleon confined to the x axis and interacting with a second nucleon on the x axis. (b) The potential energy of one nucleon in the proximity of four others fluctuates rapidly. Nevertheless, its effect is well approximated by the smoothed average, shown as a dashed curve.

*The IPA is often called the independent-particle *model*, abbreviated IPM, or the mean-field approximation.



Section 16.5 • The IPA Potential Energy for Nucleons 13

**FIGURE 16.6**

The nuclear IPA potential well, $U(r)$, seen by one nucleon under the influence of the other $A - 1$ nucleons in a nucleus. The force range R can be defined as the radius at which the potential energy is half its maximum depth and is found to be somewhat larger than the radius of the nucleus.

with their wave functions. This quantum spreading of the other particles is an important factor in the success of the IPA in describing the motion of nucleons in nuclei, as it is also for the case of electrons in an atom.

Figure 16.5 is one-dimensional, while the real nucleus is, of course, three-dimensional. We have seen that the nucleons are distributed nearly uniformly through a sphere of radius

$$R(\text{nuclear surface}) = (1.07 \text{ fm})A^{1/3} \quad (16.13)$$

Therefore, the average potential energy seen by each nucleon is found by averaging the nucleon–nucleon potential energy over this sphere. The resulting IPA potential energy, $U(r)$, is spherically symmetric and is sketched in Fig. 16.6. The radius of the potential well is given approximately by

$$R(\text{force range}) = (1.17 \text{ fm})A^{1/3} \quad (16.14)$$

Comparing this with (16.13), we see that, as one would expect, the force range is a little larger than the nuclear radius since the force can “reach out” a little beyond the distribution of nucleons. The depth of the IPA well is about 50 MeV and is approximately the same for most nuclei.

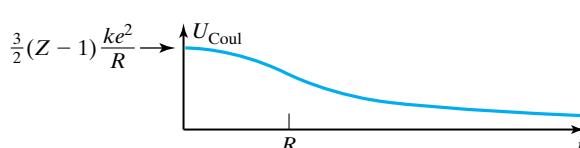
The potential-energy function in Fig. 16.6 is the averaged nuclear potential energy seen by each nucleon. Because of the charge independence of the nuclear force, this IPA potential energy is the same for neutrons as for protons. In the case of protons, however, we must add the potential energy of the electrostatic Coulomb force. To a good approximation this is the potential energy of a proton in the electric field of a uniform sphere of charge $(Z - 1)e$. Outside the nucleus ($r > R$) this is

$$U_{\text{Coul}}(r) = (Z - 1)\frac{ke^2}{r} \quad (16.15)$$

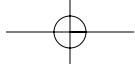
Inside the nucleus it has the shape of the inverted parabola shown in Fig. 16.7 and is maximum at the center, where it is 1.5 times its value at the nuclear surface (Problems 16.28 and 16.30):

$$U_{\text{Coul}}(0) = \frac{3}{2}(Z - 1)\frac{ke^2}{R} \quad (16.16)$$

The Coulomb potential energy (16.16) is easily calculated for any given nucleus (Problems 16.29 and 16.30). The answer varies from 1 or 2 MeV for the

**FIGURE 16.7**

The electrostatic potential energy of a proton in the electric field of the $Z - 1$ other protons in a nucleus.



14 Chapter 16 • The Structure of Atomic Nuclei

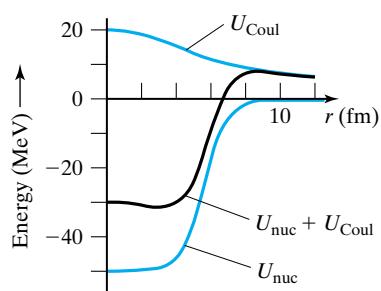


FIGURE 16.8

The IPA potential energy of any one proton in the nucleus ^{120}Sn (tin) is the sum of two parts, U_{nuc} and U_{Coul} . A neutron feels only U_{nuc} .

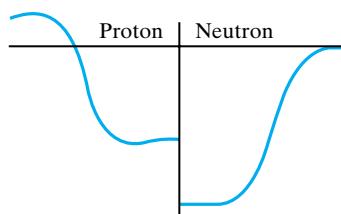


FIGURE 16.9

The IPA potential-energy functions for a proton and a neutron in a medium-mass nucleus.

lightest nuclei to about 30 MeV for the heaviest. For protons in light nuclei, it is therefore a fair approximation to ignore U_{Coul} compared to the nuclear potential energy of Fig. 16.6. For protons in heavy nuclei, this is certainly not a reasonable approximation, and we will have to use the total potential energy,

$$U(r) = U_{\text{nuc}}(r) + U_{\text{Coul}}(r)$$

This function is plotted in Fig. 16.8, which shows the IPA potential well for a proton in the middle-mass nucleus ^{120}Sn (tin).

It is often convenient to display the IPA potential-energy functions for protons and neutrons side by side. One convenient way to do this is to draw them “back to back” as in Fig. 16.9. It is important to remember that although we have drawn these two functions separately, they describe the potential energy of a proton and a neutron in one and the same nucleus.

Using the IPA potential-energy functions, one can calculate the energy levels of each nucleon in a nucleus. In this way we can construct a shell model of the nucleus, quite analogous to the atomic shell model described in Chapter 10. As we describe in Section 16.8, the nuclear shell model was developed about 1950 and is very successful in explaining many (though certainly not all) nuclear properties. Before taking up the shell model, we discuss two other ideas that depend on the existence, though not the details, of the energy levels of the wells shown in Fig. 16.9.

16.6 The Pauli Principle and the Symmetry Effect

We saw in Fig. 16.1 that there is a tendency for stable nuclei to have nearly equal numbers of protons and neutrons, especially when A is small. This tendency is often called the **symmetry effect**. Using the ideas of Section 16.5, we can now show that the symmetry effect is a simple consequence of the Pauli principle and the charge independence of nuclear forces.

Like electrons, both protons and neutrons obey the Pauli exclusion principle. That is, no two protons can occupy the same quantum state, and similarly, no two neutrons can occupy the same quantum state. It is important to understand that the Pauli principle restricts the states of two *identical* particles only; it places no restrictions on the states occupied by two particles of different type. Thus it is perfectly possible to have a proton and neutron both occupying exactly the same quantum state.

The ground state of a nucleus is found by assigning its Z protons and N neutrons to their lowest possible states consistent with the Pauli principle. To see how this works, let us first ignore the Coulomb repulsion of the protons, so that the “wells” in which the protons and neutrons move are identical. (This will be a good approximation for light nuclei, but not as good for heavier nuclei.) Figure 16.10(a) shows the two wells and the lowest few levels in each,

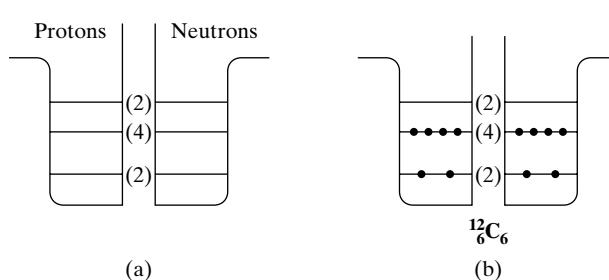
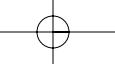
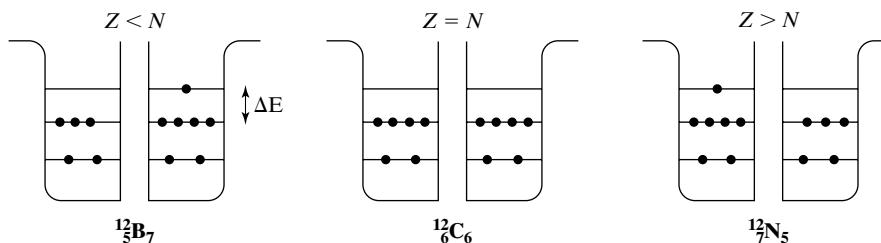


FIGURE 16.10

(a) Schematic diagram of the IPA wells and levels for protons and neutrons in a light nucleus. The two wells are essentially identical because of charge independence. The numbers in parentheses are the observed degeneracies. (b) The ground state of $^{12}\text{C}_6$ is found by putting its six protons in the lowest available proton levels and its six neutrons in the lowest available neutron levels.



Section 16.6 • The Pauli Principle and the Symmetry Effect 15

**FIGURE 16.11**

The ground states of the three isobars ^{12}B , ^{12}C , and ^{12}N . Because of the Pauli principle, the two nuclei with $Z \neq N$ have higher energy by the amount shown as ΔE .

along with their degeneracies. Since the wells are identical, the same is true of the levels. The Pauli principle requires that we place the Z protons in the lowest proton levels, with the number of protons in a level never exceeding that level's degeneracy; and similarly with the neutrons. Figure 16.10(b) shows the resulting ground state of the nucleus $^{12}_6\text{C}_6$. (The precise degeneracies of the levels do not matter for the present discussion, but we have shown their actual observed values — two for the lowest level and four for the next.)

The example shown in Fig. 16.10(b), $^{12}_6\text{C}_6$, is a nucleus with $Z = N$. Let us now consider a nucleus with the same number of nucleons, but $Z \neq N$, that is, one of the isobars of $^{12}_6\text{C}_6$. For example, consider the nucleus $^{12}_5\text{B}_7$ with five protons and seven neutrons. The ground state of $^{12}_5\text{B}_7$ is shown on the left of Fig. 16.11, which also shows the ground states of $^{12}_6\text{C}_6$ and $^{12}_7\text{N}_5$ for comparison. Comparing the ground state of $^{12}_5\text{B}_7$ with that of $^{12}_6\text{C}_6$, we see that, because of the Pauli principle, the seventh neutron in $^{12}_5\text{B}_7$ has to occupy a higher level than any of the neutrons or protons in $^{12}_6\text{C}_6$. Therefore, the nucleus $^{12}_5\text{B}_7$ has higher total energy than $^{12}_6\text{C}_6$.

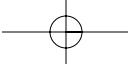
If we look on the other side of $^{12}_6\text{C}_6$ at $^{12}_7\text{N}_5$, we find a similar situation. The seventh proton of $^{12}_7\text{N}_5$ must occupy a higher level than any of the nucleons in $^{12}_6\text{C}_6$, and $^{12}_7\text{N}_5$ has higher total energy than $^{12}_6\text{C}_6$. If we move still farther away from $Z = N$ (to $^{16}\text{O}_8$, for example), the total energy has to be even higher.

Similar arguments can be applied to any set of isobars. Based on these arguments (in which we have so far ignored the Coulomb energy of the protons), we can state the following general conclusion: Among any set of isobars (nuclei with the same total number of nucleons) the nucleus (or nuclei) with Z closest to N will have the lowest total energy.

To understand what this conclusion implies, we need to anticipate a property of radioactive decay, which we discuss in Chapter 17. If a nucleus has appreciably more energy than a neighboring isobar, the nucleus with greater energy is unstable and will eventually convert itself into the isobar with lower energy. The process by which this occurs is called β decay and is discussed in Section 17.4, but for now the important point is this: Any nucleus with Z much different from N is unstable and will convert itself into a neighboring isobar with Z closer to N and hence lower total energy. Therefore, the stable nuclei, and hence the nuclei normally found to occur naturally, should all have $Z \approx N$.

In discussing the symmetry effect, we have so far ignored the electrostatic energy of the protons. This is easily taken into account: As discussed in Section 16.5, the Coulomb energy simply raises the proton well relative to the neutron well. This is shown in Fig. 16.12(a), which shows the two wells and a schematic set of energy levels. If the neutron well is appreciably lower than the proton well, there may be several neutron levels below any of the proton levels. In this case the most stable isobar will have more neutrons than protons, as indicated schematically in the figure.

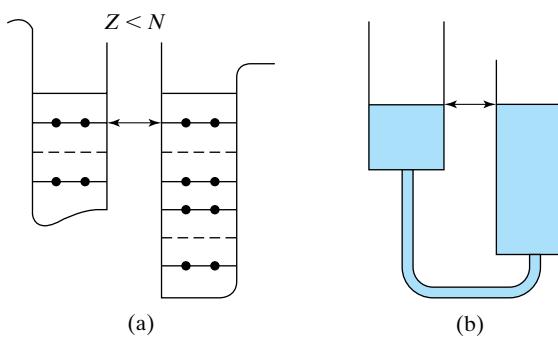
For light nuclei the difference between the proton and neutron wells is small, and our original conclusion that stable nuclei should have $Z \approx N$ is



16 Chapter 16 • The Structure of Atomic Nuclei

FIGURE 16.12

(a) In nuclei with many protons the Coulomb repulsion pushes the proton well appreciably higher than the neutron well. For a given number of nucleons, the lowest energy is obtained when the highest occupied proton and neutron levels have the same energy (as indicated by the arrow). This requires that Z be somewhat less than N . (b) The situation is closely analogous to a pair of buckets of water, whose total energy is lowest when the top surfaces are level and the higher bucket has less water. A pipe between the buckets lets the water levels adjust to this state of lowest energy, whatever the initial configuration. In nuclei the corresponding adjustment occurs through radioactive decay.



unaltered. However, as we move to heavier nuclei, the Coulomb energy steadily rises and the ratio N/Z should increase slowly, as is confirmed by the data in Fig. 16.1.

It is perhaps worth reiterating the important role of the Pauli principle in the symmetry effect. In Fig. 16.12(a) it is clear that without the Pauli principle, the state of lowest energy would consist in all (or almost all) the nucleons being neutrons, all occupying the lowest neutron level. Under these conditions the majority of elements that make up our world would not exist.

16.7 The Semiempirical Binding-Energy Formula

The binding energy B of a nucleus is the total energy needed to separate completely its A individual nucleons. Binding energies of different nuclei are a useful measure of their relative stabilities as we will see later. In this section we discuss how the binding energy depends on the nuclear parameters Z , N , and $A = Z + N$. We start with a brief discussion of the measurement of binding energies.

As mentioned in Section 3.5, nuclear binding energies are large enough to cause an observable difference (of order 1%) between the mass of a nucleus and the sum of the separate masses of its Z protons and N neutrons:

$$m_{\text{nuc}} = Zm_p + Nm_n - \frac{B}{c^2} \quad (16.17)$$

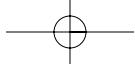
where m_{nuc} denotes the mass of the nucleus concerned. Given accurate measurements of the masses involved, we can use this relation to find B as

$$B = (Zm_p + Nm_n - m_{\text{nuc}})c^2 \quad (16.18)$$

Most mass tabulations, including that in Appendix D, give the mass m_{atom} of the atom, rather than that of the bare nucleus, since it is m_{atom} that is usually measured. The mass m_{atom} is equal to m_{nuc} plus the mass of the Z atomic electrons*

$$m_{\text{atom}} = m_{\text{nuc}} + Zm_e$$

*There is, of course, a small correction due to the binding energy of the electrons. Since average electron binding energies are at most several keV, while nuclear binding energies are several MeV, this correction is almost always negligible.



Section 16.7 • The Semiempirical Binding-Energy Formula 17

Now, we can add Zm_e to m_{nuc} in (16.18) if we also add Zm_e to the term Zm_p . This replaces Zm_p by

$$Z(m_p + m_e) = Zm_H$$

where m_H is the mass of the hydrogen atom (${}^1\text{H}$). This means that we can rewrite (16.18) as

$$B = (Zm_H + Nm_n - m_{\text{atom}})c^2 \quad (16.19)$$

and work entirely with atomic masses.

Example 16.3

Use the atomic masses in Appendix D to find the binding energy of the nucleus ${}^{35}\text{Cl}_{18}$.

According to (16.19), the required binding energy is

$$B = [17m_H + 18m_n - m_{\text{atom}}({}^{35}\text{Cl})]c^2$$

or, with the data from Appendix D (rounded to five significant figures),

$$\begin{aligned} B &= [17(1.0078 \text{ u}) + 18(1.0087 \text{ u}) - (34.969 \text{ u})]c^2 \\ &= [35.289 \text{ u} - 34.969 \text{ u}]c^2 \end{aligned} \quad (16.20)$$

$$\begin{aligned} &= (0.329 \text{ u}) \times \frac{931.5 \text{ MeV}/c^2}{1 \text{ u}} \\ &= 298 \text{ MeV} \end{aligned} \quad (16.21)$$

Notice in (16.20) that the mass of the separate constituents is *greater* than that of ${}^{35}\text{Cl}$ confirming that we must *add* energy to pull the nucleus apart. Notice also that (16.20) gives B as the difference of two masses that are very nearly equal. This difference is always of order 1% or less. Thus, we must know the masses to at least two significant figures more than are required in our answer. Here we needed five significant figures in the masses to get three in B .

As this example shows, one can find the binding energy of a nucleus if one knows the corresponding atomic mass accurately. Using a mass spectrometer, it is possible to measure atomic masses very accurately (see Section 16.9), and this is therefore one of the best ways to find nuclear binding energies.

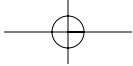
The direct measurement of binding energies, by complete separation of a nucleus, is seldom possible. However, it is often possible to find the binding energy of one nucleus in terms of the known binding energy of some other nucleus. For example, one can measure the energy needed to separate one neutron from a nucleus. This is called the *neutron separation energy*,

$$S_n(Z, N) = \text{energy to separate one neutron from nucleus } (Z, N)$$

It is easy to see that the binding energy of the nucleus (Z, N) is

$$B(Z, N) = S_n(Z, N) + B(Z, N - 1) \quad (16.22)$$

that is, the energy to dismember (Z, N) completely equals the energy to pull off one neutron *plus* the energy to dismember the remaining nucleus $(Z, N - 1)$.



18 Chapter 16 • The Structure of Atomic Nuclei

Thus, if the binding energy of $(Z, N - 1)$ is known, we can use (16.22) to find the binding energy of (Z, N) or vice versa.

Example 16.4

The binding energy of $^{16}\text{O}_8$ is known to be 127.6 MeV. It is found that 4.2 MeV is needed to separate a neutron from $^{17}\text{O}_9$. What is the binding energy of $^{17}\text{O}_9$?

According to (16.22), the required binding energy is

$$\begin{aligned} B(^{17}\text{O}) &= S_n(^{17}\text{O}) + B(^{16}\text{O}) \\ &= (4.2 \text{ MeV}) + (127.6 \text{ MeV}) = 131.8 \text{ MeV} \end{aligned}$$

In the remainder of this section we derive a formula that gives the binding energy of nuclei as a function of the numbers Z, N , and $A = Z + N$. All of the terms in this formula have at least partial theoretical justifications, but the values of the coefficients that appear are generally found by fitting the formula to the measured binding energies. For this reason the formula is called the **semiempirical binding-energy formula**, or the **Weizsäcker semiempirical formula**, after its inventor. If we insert the formula for B into the equation $m_{\text{nuc}} = Zm_p + Nm_n - B/c^2$, the resulting equation is called the semiempirical mass formula.

To an excellent approximation, the binding energy of a nucleus is found to be proportional to the number, A , of its nucleons. This result is easily understood by the following classical argument: We have seen that the nuclear force has a range of about 2 fm, which is smaller than most nuclear sizes. For most nuclei, then, each nucleon is bound to only a fraction of the other nucleons, namely, those neighbors within a sphere of radius about 2 fm, as shown in Fig. 16.13. We have also seen that the density of nucleons is about the same inside all nuclei. It follows that the number of neighbors to which each nucleon is bound is approximately the same for all nuclei. That is, the average binding energy of each nucleon is the same in all nuclei, so that the *total* binding energy is proportional to A . Since A is proportional to the volume of the nucleus, this approximation is called the *volume term* and is written

$$B \approx B_{\text{vol}} = a_{\text{vol}}A \quad (16.23)$$

where a_{vol} is a positive constant.

There is an immediate correction to the approximation (16.23). Those nucleons near the nuclear surface have fewer neighbors to attract them (Fig. 16.13) and are less tightly bound than those in the interior. The number of these surface nucleons is proportional to the surface area of the nucleus, $4\pi R^2 = 4\pi R_0^2 A^{2/3}$. Accordingly, (16.23) must be reduced by an amount proportional to $A^{2/3}$, and our next approximation is

$$B \approx B_{\text{vol}} + B_{\text{surf}}$$

where the surface correction B_{surf} is negative and has the form

$$B_{\text{surf}} = -a_{\text{surf}} A^{2/3} \quad (16.24)$$

with a_{surf} a positive constant.

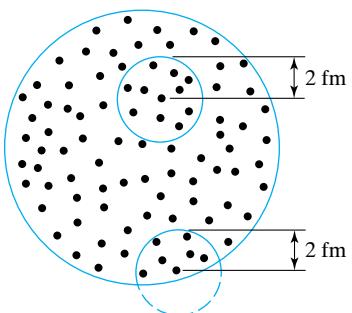
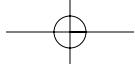


FIGURE 16.13

A nucleon in a nucleus is bound only to those neighbors within a sphere of radius about 2 fm (upper small sphere); thus all interior nucleons are bound to about the same number of neighbors. A nucleon near the surface is bound to fewer neighbors (lower small sphere), and this slightly reduces the total binding energy.

Section 16.7 • The Semiempirical Binding-Energy Formula **19**

A further correction is needed because of the electrostatic repulsion of the protons. This reduces the binding energy by an amount equal to the potential energy of the total nuclear charge. The potential energy of a uniform sphere of charge Ze is (Problem 16.39)

$$U_{\text{Coul}} = \frac{3}{5} \frac{k(Ze)^2}{R} \quad (16.25)$$

Since $R = R_0 A^{1/3}$, this implies a correction of the form

$$B_{\text{Coul}} = -a_{\text{Coul}} \frac{Z^2}{A^{1/3}} \quad (16.26)$$

where

$$a_{\text{Coul}} \approx \frac{3}{5} \frac{ke^2}{R_0}$$

Since the nucleus is not precisely a uniform sphere of charge, this is not the exact value of a_{Coul} .

The next correction arises from the symmetry effect discussed in Section 16.6, where we saw that for any set of isobars, the nucleus with $Z = N$ has highest binding energy. Since B gets less as $Z - N$ increases in either direction, this suggests a negative term proportional to $(Z - N)^2$. The observed binding energies of all nuclei are best fitted by a correction of the form

$$B_{\text{sym}} = -a_{\text{sym}} \frac{(Z - N)^2}{A} \quad (16.27)$$

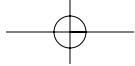
The factor of A in the denominator reflects that for larger nuclei, the level spacing is smaller and the difference of energy between neighboring isobars (shown as ΔE in Fig. 16.11) is therefore smaller.

The final term in our formula for the binding energy reflects a phenomenon that we will discuss in more detail in the next section. We mentioned in Section 16.2 that the majority of stable nuclei (148, in fact) have both Z and N even; there are 100 nuclei with Z even but N odd or vice versa; and there are only 4 in which both Z and N are odd. This preference for Z and N to be even is explained by a property of the nuclear force called the **pairing effect**. For now, we simply state that this effect gives a correction to the binding energy with the form

$$B_{\text{pair}} = \varepsilon \frac{a_{\text{pair}}}{A^{1/2}} \quad (16.28)$$

where a_{pair} is some constant and

$$\varepsilon = \begin{cases} 1 & \text{if } Z \text{ and } N \text{ are both even} \\ 0 & \text{if } Z \text{ or } N \text{ is even and the other is odd} \\ -1 & \text{if } Z \text{ and } N \text{ are both odd} \end{cases} \quad (16.29)$$



20 Chapter 16 • The Structure of Atomic Nuclei

Putting all five of these contributions together, we obtain the **semiempirical binding-energy formula**

$$\begin{aligned} B &= B_{\text{vol}} + B_{\text{surf}} + B_{\text{Coul}} + B_{\text{sym}} + B_{\text{pair}} \\ &= a_{\text{vol}}A - a_{\text{surf}}A^{2/3} - a_{\text{Coul}}\frac{Z^2}{A^{1/3}} - a_{\text{sym}}\frac{(Z - N)^2}{A} + \epsilon\frac{a_{\text{pair}}}{A^{1/2}} \end{aligned} \quad (16.30)$$

where ϵ is defined in (16.29). The five coefficients are generally adjusted to fit all the measured binding energies as well as possible. One set of values that gives a good fit is as follows:

$$\left. \begin{array}{ll} a_{\text{vol}} = 15.75 \text{ MeV} & a_{\text{surf}} = 17.8 \text{ MeV} \\ a_{\text{Coul}} = 0.711 \text{ MeV} & a_{\text{sym}} = 23.7 \text{ MeV} \\ a_{\text{pair}} = 11.2 \text{ MeV} & \end{array} \right\} \quad (16.31)$$

Because the dominant behavior of B is $B \propto A$, it is usual to discuss B/A , the **binding energy per nucleon**. Figure 16.14 shows the measured value of B/A for a large number of nuclei. The smooth curve is the value of B/A predicted by (16.30) and is seen to give an excellent fit for all but the lightest nuclei (those with A below about 20). It is not surprising that the smooth function (16.30) does not work well for the light nuclei, in which the quantization of energy levels is proportionately more important. Nevertheless, even when A is small, the *general* trend of B/A is correctly given by (16.30).

The most obvious feature of Fig. 16.14 is that B/A is close to 8 MeV for almost all nuclei, with the maximum value of 8.8 MeV for iron (Fe). Beyond iron, B/A slopes gently down to about 7.6 MeV for ^{238}U ; this decrease is due mainly to the increasing importance of the Coulomb repulsion of the protons. On the other side of ^{56}Fe , when A decreases below about 20, B/A falls rapidly to zero for ^1H (which has no binding energy); this decrease occurs because almost all nucleons in a small nucleus are close to the surface and the negative surface correction is proportionately large.

The lower values of B/A when A is large imply that energy is released when a heavy nucleus splits, or *fissions*, into two lighter nuclei. The lower values

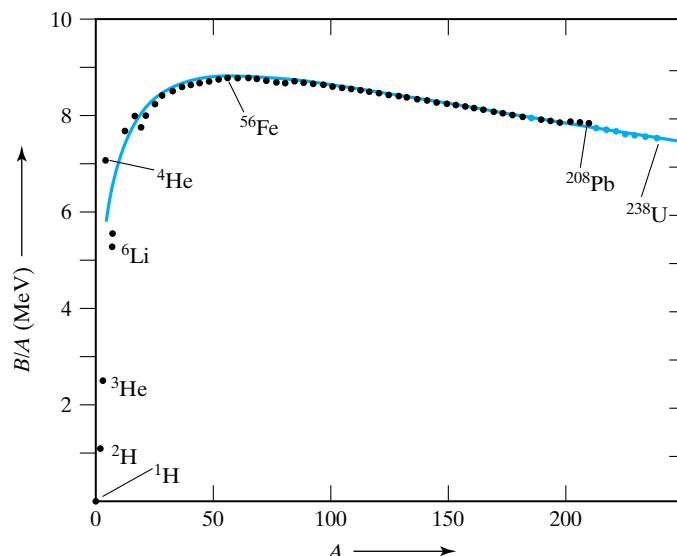


FIGURE 16.14

The binding energy per nucleon, B/A , as a function of A . The points are measured values. (Black dots indicate stable nuclei; colored, unstable.) The curve is the prediction of the binding-energy formula (16.30). Values of B/A are shown for the most stable isobar of each given A . For $A > 7$, only even-even nuclei are shown, and for clarity, only half the data are included. The curve was calculated from (16.30), assuming that both Z and N are even.

of B/A when A is very small imply that energy is also released when two light nuclei fuse to form one heavier nucleus. Nuclear fission is the source of energy in nuclear power stations, while fusion provides the energy of the stars, including our own sun. Both fission and fusion are used in nuclear weapons. Clearly, the behavior of B/A , as shown in Fig. 16.14, is of great practical significance.

Example 16.5

Use the binding-energy formula (16.30) to predict the binding energy of ^{35}Cl , and compare the result with the measured value found in (16.21).

Inserting the values $A = 35$, $Z = 17$, $N = 18$ and the coefficients (16.31) into the formula (16.30), we find that

$$\begin{aligned} B &= a_{\text{vol}}A - a_{\text{surf}}A^{2/3} - a_{\text{Coul}}\frac{Z^2}{A^{1/3}} - a_{\text{sym}}\frac{(Z - N)^2}{A} + \epsilon\frac{a_{\text{pair}}}{A^{1/2}} \\ &= (551.2 - 190.5 - 62.8 - 0.7 + 0) \text{ MeV} \\ &= 297 \text{ MeV} \end{aligned}$$

which agrees very well with the measured value of 298 MeV. Notice that the pairing term is zero in this example because Z is odd and N is even.

16.8 The Shell Model*

*The shell model is an important part of our understanding of nuclei and is a beautiful application of many of the ideas of the atomic shell model. Nevertheless, we will not be using the ideas of this section again, and you could omit it without loss of continuity.

The semiempirical binding-energy formula gives a remarkably accurate picture of the general trends of nuclear binding energies. Nevertheless, it is a smoothed average picture and shows none of the fluctuations that one would expect if nucleons, like atomic electrons, occupy energy levels that fall into distinct shells. Recall that atomic binding energies fluctuate periodically with Z , being greatest when the electrons exactly fill a closed shell and dropping steeply just beyond each closed shell. In this section we discuss the corresponding shell properties of nuclei.

The Magic Numbers

The total numbers of electrons in the closed-shell atoms are known to be

atomic closed-shell numbers:

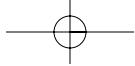
$$2, 10, 18, 36, 54, 86 \quad (16.32)$$

By the late 1940s, substantial evidence had accumulated that nuclei have analogous closed-shell numbers which lead to unusual stability, and that these **magic numbers**, as they came to be called, were

nuclear magic numbers:

$$2, 8, 20, 28, 50, 82, 126 \quad (16.33)$$

A nucleus is especially stable if Z or N is equal to one of these magic numbers, and even more so if both are. The magic numbers are the same for protons as for neutrons, as one would expect from the charge independence of nuclear

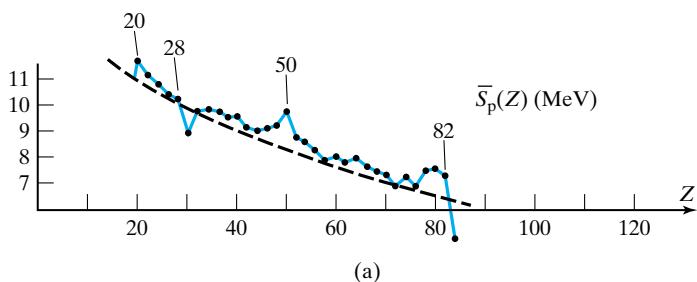


22 Chapter 16 • The Structure of Atomic Nuclei

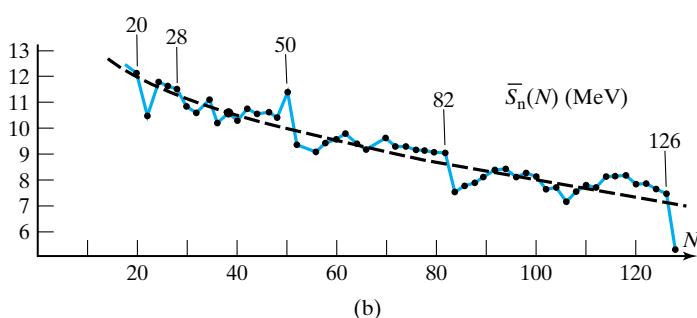
forces.* On the other hand, the nuclear magic numbers are quite different from the atomic closed-shell numbers (16.32), reflecting the great difference between the dominant forces in nuclei and atoms.

The evidence for the magic numbers (16.33) consists of many observations, few of which are conclusive by themselves. Nuclei in which Z is magic tend to have larger numbers of stable isotopes than average. For example, tin, with Z equal to the magic number 50, has 10 stable isotopes, more than any other element. Similarly, nuclei in which N is magic tend to have more stable isotones than average; $N = 82$ holds the record, with 7 stable isotones. The binding energy per nucleon, B/A , in each of the “doubly magic” nuclei $^2\text{He}_2$, $^{80}\text{O}_8$, $^{20}\text{Ca}_{20}$, $^{20}\text{Ca}_{28}$, and $^{82}\text{Pb}_{126}$ is markedly higher than the prediction of the binding-energy formula. (This is easily discernible in the case of $^2\text{He}_2$ on the graph of B/A in Fig. 16.14.)

Perhaps the best evidence for the magic numbers (16.33) concerns the separation energies S_p and S_n . These are the energies required to remove one proton or one neutron from a nucleus and are analogous to the ionization energy of an atom. Recall that a plot of atomic ionization energy against number of electrons (Fig. 10.11) shows a maximum at each of the closed shells, with an abrupt drop immediately thereafter. The nuclear separation energies depend on two variables, Z and N , and are harder to plot. One simple procedure is to eliminate one of these variables by an appropriate averaging. If we consider, for example, the proton separation energy $\bar{S}_p(Z, N)$, then for each value of Z we can compute the average $\bar{S}_p(Z)$ for all stable isotopes with this Z . A plot of $\bar{S}_p(Z)$ against Z is shown in Fig. 16.15(a) and, as expected, has pronounced drops immediately after each of the magic numbers indicated. (We have not included data for $Z < 20$, since our averaging procedure obscures shell effects for very light nuclei.) The dashed curve in Fig. 16.15(a) shows the proton separation energy



(a)



(b)

FIGURE 16.15

(a) The proton separation energy $\bar{S}_p(Z)$ averaged over all stable isotopes for each Z . (b) The neutron separation energy $\bar{S}_n(N)$ averaged over all stable isotones for each N . The dashed curves show the predictions of the binding-energy formula (16.30). Only even values of Z were included in (a) and only even N in (b).

*Note, however, that nuclei with $Z = 126$ have not been observed. Therefore, there is no experimental evidence for a proton magic number in this region.

Section 16.8 • The Shell Model 23

predicted by the binding-energy formula. This reproduces the general trend very well but entirely misses the fluctuations associated with the magic numbers.

Figure 16.15(b) shows the average neutron separation energies $\bar{S}_n(N)$, each averaged over all stable isotones of a given N . This shows abrupt drops beyond the same magic numbers, 20, 28, 50, and 82, as in Fig. 16.15(a), and also at $N = 126$.

Origin of the Magic Numbers

Having established the existence of the nuclear magic numbers (16.33), physicists naturally hoped to explain them in terms of nucleon energy levels and the grouping of those levels into energy shells. The IPA potential energy felt by any one proton or neutron was described in Section 16.5 and sketched in Fig. 16.6. The energy levels for this well can be calculated numerically, and the ordering of the levels is shown in Fig. 16.16.

The levels shown in Fig. 16.16 are labeled according to the scheme normally used in nuclear physics. This scheme is nearly, although not quite, the same as that used in atomic physics. Just as in atomic physics, each level is identified by two quantum numbers, n and l . The number l identifies the magnitude of the nucleon's orbital angular momentum, $L = \sqrt{l(l+1)}\hbar$. The possible values $l = 0, 1, 2, 3, 4, \dots$ are indicated by the code letters s, p, d, f, g, \dots . For each value of l , there are $2l + 1$ possible values of $L_z = m\hbar$ (with $m = l, l-1, \dots, -l$) and two possible values of $S_z = \pm \frac{1}{2}\hbar$. The degeneracy of each level is therefore $2(2l+1)$. Evidently, all of the parameters associated with angular momentum are the same as in atomic physics. On the other hand, nuclear physicists define the quantum number n so that the levels of a given l are simply numbered $n = 1, 2, 3, \dots$. Thus the levels with $l = 1$ are labeled $1p, 2p, 3p, \dots$ in order of increasing energy; the levels with $l = 2$ are $1d, 2d, 3d, \dots$, and so on.*

The levels (or subshells) shown in Fig. 16.16 are grouped into well-defined energy shells, and the total numbers of protons or neutrons needed to produce closed shells are indicated on the right. The first three of these numbers agree exactly with the observed magic numbers, but beyond this, all of the predicted closed-shell numbers turn out to be wrong.

At first it would seem natural to assume that a different shape for the IPA potential well might produce the correct magic numbers. In fact, however, no reasonable well gives the observed numbers. The correct explanation was discovered independently in 1949 by Maria Goeppert-Mayer and by Johannes Jensen and coworkers. It turns out that the nuclear force depends strongly on the relative orientation of each nucleon's spin and orbital angular momentum. Specifically, the IPA potential well of Fig. 16.6 is the correct potential-energy function for a nucleon with $l = 0$, but if $l \neq 0$, the nucleon has an additional potential energy that depends on the orbital angular momentum \mathbf{L} and its orientation relative to the spin \mathbf{S} . This additional energy ΔU is found to have the form

$$\Delta U \propto \mathbf{S} \cdot \mathbf{L} \cos \theta$$

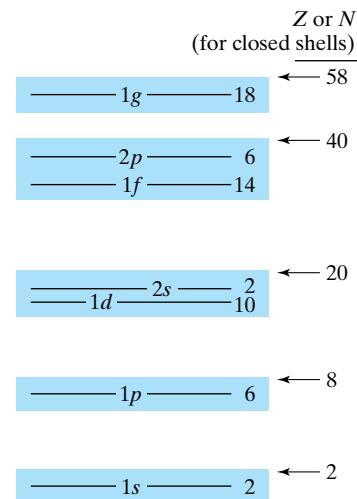
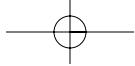


FIGURE 16.16

The (incorrect) filling order of the levels for a proton or neutron, calculated for the nuclear potential well of Fig. 16.6. The number to the right of each level is its degeneracy. The levels fall into distinct energy shells, but the closed-shell numbers do not agree with the observed magic numbers (16.33). It follows that these levels are not correct.

*Recall that in atomic physics the lowest p state is called $2p$, the lowest d state $3d$, and so on. This scheme is very convenient in the hydrogen atom (where it means that all states of given n have the same energy) but has no particular merit in nuclear physics.



24 Chapter 16 • The Structure of Atomic Nuclei

Maria Goeppert-Mayer
(1906–1972, German-US)



After getting her PhD under Max Born at Gottingen in 1933, Goeppert-Mayer moved to the United States. During most of her career, nepotism rules at universities where her husband worked prevented Mayer from holding official positions. In 1949, while at the University of Chicago, she proposed the nuclear shell model, for which she shared the 1963 Nobel Prize in physics with Jensen. Although Goeppert-Mayer was an active member of the Chicago faculty — teaching courses, advising graduate students, and doing groundbreaking research — she had no official appointment and no salary. In 1960 she and her husband moved to the University of California, San Diego, where she was appointed professor of physics, her first full-time paid position. The picture shows her talking with Enrico Fermi at a summer school at the University of Michigan in the 1930s.

Johannes Jensen
(1907–1973, German)



Jensen was a professor first at Hamburg and then at Heidelberg. In 1949 he and Goeppert-Mayer independently proposed the nuclear shell model. Together they wrote a book on the model in 1955, and they shared the Nobel Prize in physics in 1963.

where θ is the angle between \mathbf{S} and \mathbf{L} ; that is,

$$\Delta U \propto \mathbf{S} \cdot \mathbf{L} \quad (16.34)$$

For this reason, the energy ΔU is called the **spin-orbit energy**.

Since the spin-orbit energy of each nucleon is proportional to $\mathbf{S} \cdot \mathbf{L}$, each energy level of Fig. 16.16 (with $l \neq 0$) is split into two levels according to the two possible orientations of \mathbf{S} relative to \mathbf{L} . (Recall that \mathbf{S} has just two possible orientations relative to any direction.) A spin-orbit splitting of this kind is familiar in atomic physics, where it produces the so-called fine structure of atomic spectra. As described in Section 9.7, an orbiting electron sees a magnetic field $\mathbf{B} \propto \mathbf{L}$; since the electron has a magnetic moment $\boldsymbol{\mu} \propto \mathbf{S}$, this leads to a magnetic energy $-\boldsymbol{\mu} \cdot \mathbf{B} \propto \mathbf{S} \cdot \mathbf{L}$. However, the nuclear spin-orbit splitting is *not* a magnetic effect. Rather, it is a new property of the nuclear force. The nuclear splitting is much larger than the atomic effect and is in the opposite direction: For an electron in an atom, the state with \mathbf{S} and \mathbf{L} parallel has slightly *higher* energy; for a nucleon in a nucleus, the corresponding state has much *lower* energy. The first clear evidence for this nuclear spin-orbit energy came from studies of nuclear shell structure, but scattering experiments with separate spin-up and spin-down beams of nucleons have since confirmed its existence and properties.

The effect of the spin-orbit energy on the levels of Fig. 16.16 is shown in Fig. 16.17, which shows the resulting levels (or subshells) with their degeneracies and their grouping into energy shells. As one would expect, the effect of the spin-orbit energy increases with increasing l : The $1s$ level, like all s levels, is not split at all; the $1p$ level, like all p levels (and also d levels), is split, but the resulting two levels are still in the same shell; for the $1f$ level and all levels with $l \geq 3$, the splitting is so large that the resulting two levels belong to different shells. It can be seen that as a result of these splittings, the closed-shell numbers agree perfectly with the observed magic numbers.

To understand the labeling and degeneracies of the levels in Fig. 16.17, we must discuss briefly the total angular momentum \mathbf{J} of a nucleon,

$$\mathbf{J} = \mathbf{L} + \mathbf{S}$$

The magnitude of \mathbf{L} is $L = \sqrt{l(l+1)\hbar}$ with $l = 0, 1, 2, \dots$, and that of \mathbf{S} is $S = \sqrt{s(s+1)\hbar}$, with s always equal to $\frac{1}{2}$. It is found that the magnitude of \mathbf{J} is given by a similar formula,

$$J = \sqrt{j(j+1)\hbar} \quad (16.35)$$

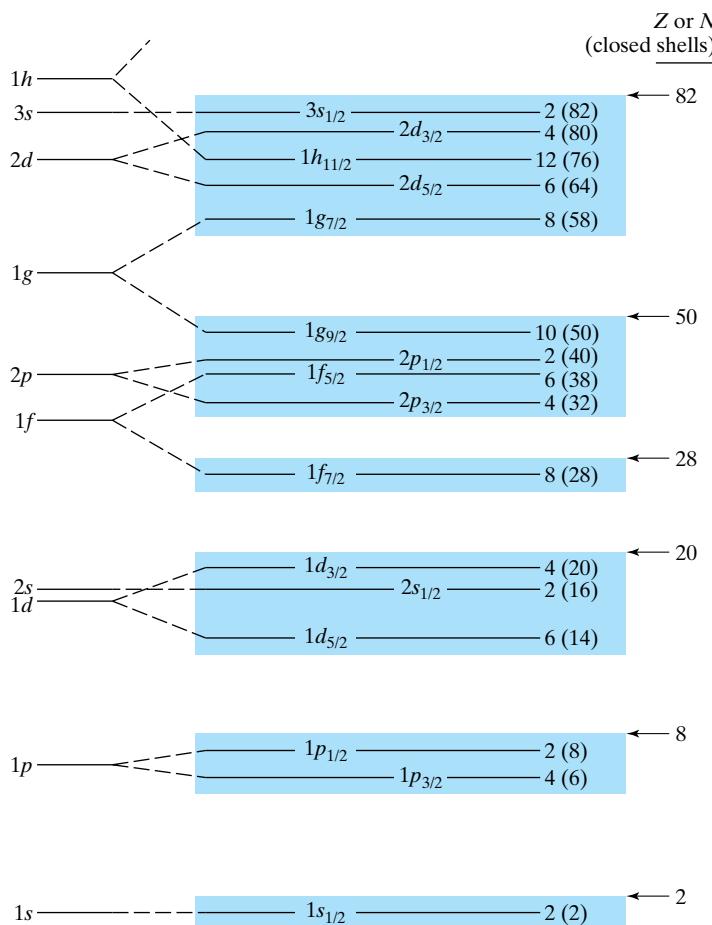
If its orbital angular momentum is zero ($l = 0$), the nucleon's total angular momentum is just the spin, and hence $j = \frac{1}{2}$. If $l \neq 0$, it is found that there are two possible values of j , depending on the orientation of \mathbf{S} relative to \mathbf{L} : If \mathbf{S} is parallel to \mathbf{L} , then

$$j = l + \frac{1}{2} \quad (\mathbf{S} \text{ and } \mathbf{L} \text{ parallel}) \quad (16.36)$$

but if \mathbf{S} is antiparallel to \mathbf{L} , then

$$j = l - \frac{1}{2} \quad (\mathbf{S} \text{ and } \mathbf{L} \text{ antiparallel}) \quad (16.37)$$

Section 16.8 • The Shell Model 25

**FIGURE 16.17**

Filling order of the levels through Z or $N = 82$ for a single nucleon in a nucleus, including the spin-orbit energy proposed by Goeppert-Mayer and Jensen. The levels on the left are the corresponding levels in the absence of the spin-orbit energy. The numbers to the right of each level are the level's degeneracy and (in parentheses) the running total of protons or neutrons needed to fill through that level. On the far right are the closed-shell numbers, which agree perfectly with the observed magic numbers (16.33). The ordering of certain nearby levels is ambiguous (just as it is in atoms) and can be different for protons and neutrons. Beyond 82, where the proton well is strongly distorted by Coulomb repulsion, the level orderings for protons and neutrons are significantly different.

In Fig. 16.17 the value of j in each level is shown by a subscript following the code letter for l . Thus the original $1p$ level, shown on the left, is really two levels with

$$\begin{aligned} j &= l \pm \frac{1}{2} = 1 \pm \frac{1}{2} \\ &= \frac{3}{2} \quad \text{or} \quad \frac{1}{2} \end{aligned}$$

These are shown on the right of Fig. 16.17 as $1p_{3/2}$ and $1p_{1/2}$, with $1p_{3/2}$ somewhat lower in energy.

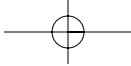
For each magnitude of the total angular momentum, there are several possible orientations of \mathbf{J} , all with the same energy. These correspond to the different possible values of J_z , which are given by

$$J_z = m_j \hbar$$

where m_j can have any of the $2j + 1$ values

$$m_j = j, j - 1, \dots, -j$$

Thus, each of the levels of Fig. 16.17 has a degeneracy of $(2j + 1)$, as indicated to the right of each level.



26 Chapter 16 • The Structure of Atomic Nuclei

Example 16.6

Find the degeneracies of the $1p_{3/2}$ and $1p_{1/2}$ levels. Show that the sum of these is the same as the total degeneracy of the original $1p$ level on the left of Fig. 16.17.

The required degeneracies are

$$1p_{3/2}: \text{degeneracy} = 2j + 1 = \left(2 \times \frac{3}{2}\right) + 1 = 4 \quad (16.38)$$

and

$$1p_{1/2}: \text{degeneracy} = 2j + 1 = \left(2 \times \frac{1}{2}\right) + 1 = 2 \quad (16.39)$$

In the absence of the $\mathbf{S} \cdot \mathbf{L}$ energy, there would have been a single $1p$ level whose degeneracy would have been $2(2l + 1) = 6$ (that is, two orientations of the spin times three of \mathbf{L}). This is equal to the sum of (16.38) and (16.39).

It can be shown quite generally (Problem 16.55) that for any given values of n and l , the sum of the degeneracies of the two corresponding levels (with $j = l \pm \frac{1}{2}$) is equal to the total degeneracy of the single level, nl , that would have existed in the absence of the $\mathbf{S} \cdot \mathbf{L}$ energy. Thus, the effect of the $\mathbf{S} \cdot \mathbf{L}$ term is simply to redistribute the degeneracies so as to produce the observed magic numbers.

Angular Momenta of Nuclei

In Section 10.7 we saw that one can predict the angular momentum of many atoms when one knows the order in which the electron energy levels are filled. We can now apply similar reasoning to nuclei. We first note that any filled level of given j contributes zero total angular momentum. (This is because for every nucleon with $J_z = m_j\hbar$, there is a second with $J_z = -m_j\hbar$.) This implies that any nucleus in which all of the occupied levels are closed (no partially filled levels) should have zero total angular momentum. That is,

$$J_{\text{tot}} = \sqrt{j_{\text{tot}}(j_{\text{tot}} + 1)}\hbar = 0$$

in this case. For example, in ${}^6\text{C}_6$ the protons and neutrons both completely fill their $1s_{1/2}$ and $1p_{3/2}$ levels. Therefore, ${}^6\text{C}_6$ should have, and does have, zero total angular momentum, $j_{\text{tot}} = 0$. Referring to Fig. 16.17, we can predict in the same way that the closed-subshell nuclei ${}^8\text{O}_8$, ${}^{14}\text{Si}_{16}$, and ${}^{20}\text{Ca}_{28}$ should all have $j_{\text{tot}} = 0$, and in every case this is found to be correct.

Let us consider next a nucleus with a single nucleon outside otherwise filled subshells. For example, in ${}^8\text{O}_9$ the eight protons and eight of the nine neutrons fill the levels $1s_{1/2}$, $1p_{3/2}$, and $1p_{1/2}$, but the ninth neutron must occupy the $1d_{5/2}$ level all by itself. Since the closed levels all have zero angular momentum, the angular momentum of the whole nucleus is just that of the final, odd neutron. Thus we predict that ${}^8\text{O}_9$ should have $j_{\text{tot}} = 5/2$, and this proves to be correct.

Throughout this section we have been discussing the ground states of nuclei, but we can also predict the properties of some excited states. For example, we just saw that the ground state of ${}^8\text{O}_9$ has its last neutron in the $1d_{5/2}$ level. In Fig. 16.17 we see that just above the $1d_{5/2}$ level is the $2s_{1/2}$. Hence we would expect that the first excited state of ${}^8\text{O}_9$ would be obtained by lifting the last neutron to the $2s_{1/2}$ level, and should have $j_{\text{tot}} = 1/2$. This is observed to be the case.

Pairing Energy

Let us next consider a nucleus with two identical nucleons (two protons or two neutrons) outside otherwise filled subshells. We could, for example, consider ${}^8\text{O}_{10}$ with two neutrons in the $1d_{5/2}$ level. The filled levels all have zero angular momentum, but since each of the last two neutrons has $j = 5/2$ there are several possible values of j_{tot} , depending on the relative orientation of the two separate angular momenta. However, it turns out that the state in which the two angular momenta are antiparallel ($j_{\text{tot}} = 0$) has significantly lower energy. Therefore, the ground state of ${}^8\text{O}_{10}$ has $j_{\text{tot}} = 0$.

It is found quite generally that any pair of identical nucleons in the same level has lowest energy when their total angular momentum is zero. The extra binding energy in the $j_{\text{tot}} = 0$ state is called the **pairing energy** and is typically a few MeV. This energy is the origin of the term B_{pair} , (16.28), in the sem empirical binding-energy formula.

Whenever a level is partially filled, the angular momenta of its nucleons can combine to give various different total angular momenta. However, the pairing effect implies that in the state of lowest energy the nucleons will be arranged, as far as possible, in pairs with zero angular momentum. If the number of nucleons in the level is even, this means that the state of lowest energy has $j_{\text{tot}} = 0$; if the number of nucleons is odd, the state of lowest energy has all but one nucleon arranged in $j = 0$ pairs and the total angular momentum is just that of the odd, unpaired nucleon.

These properties lead us to the following useful rules:

The ground state of any nucleus in which both Z and N are even
(an “even–even” nucleus) has zero total angular momentum. (16.40)

and

If a nucleus has A odd (that is, Z is even and N odd or vice versa),
the total angular momentum of its ground state is just that of the
last unpaired nucleon. (16.41)

Nuclei in which both Z and N are odd are harder to analyze,* but there are only four such stable nuclei in any case.

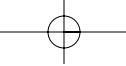
The rule (16.40) is true for all known even–even nuclei. The rule (16.41), together with the level ordering shown in Fig. 16.17, correctly predicts the angular momentum of most odd- A nuclei; the only exceptions involve nonspherical nuclei, for which the ordering of Fig. 16.17 is not always applicable. (See Problem 16.53.) The use of both rules is illustrated in the following example.

Example 16.7

What are the quantum numbers j_{tot} , for the total angular momenta of the nuclei ${}^{22}\text{Ne}$ and ${}^{43}\text{Ca}$?

Since ${}^{22}\text{Ne}$ has $Z = 10$ and $N = 12$, it is an even–even nucleus and hence has $j_{\text{tot}} = 0$. The nucleus ${}^{43}\text{Ca}$ has $Z = 20$ and $N = 23$. Since Z is even, the protons have zero total angular momentum. The total angular momentum of the neutrons is just that of the twenty-third neutron, which is in the $1f_{7/2}$ level (Fig. 16.17). Hence the nucleus ${}^{43}\text{Ca}$ should, and does, have $j_{\text{tot}} = 7/2$.

*The difficulty is to predict which relative orientation of the angular momenta of the last proton and last neutron gives the lowest energy.



28 Chapter 16 • The Structure of Atomic Nuclei

16.9 Mass Spectrometers*

* As we describe in this section, the mass spectrometer can measure atomic and nuclear masses with extraordinary precision and has become a powerful tool of chemical analysis. Nevertheless, we will not be using the ideas of this section again, and you could omit it if pressed for time.

In this section we describe mass spectrometers, the devices used to measure nuclear masses. As we saw earlier, measurements of nuclear masses contributed to our understanding of nuclear structure in two important ways. First, early measurements, which could distinguish masses differing by a few percent, revealed the existence of isotopes. Second, more precise measurements were able to determine the small relativistic shifts (about 1%) in nuclear masses due to the binding energy. The variation of binding energy with mass number, A , revealed several properties of the nuclear force, as we discussed in Section 16.7. Today, mass spectrometers can measure masses with an accuracy of order 1 part in 10^8 , which allows binding energies of typical nuclei to be calculated within 1 keV or so. Most recently, the mass spectrometer has developed into a powerful tool of chemical analysis, as we describe shortly.

In practice, it is generally the mass of an ionized atom (or molecule) that is measured in a mass spectrometer. The substance to be analyzed is introduced into an electric discharge, which causes some atoms to lose one electron and become positive ions of charge e . In the arrangement shown in Fig. 16.18, the ions are next accelerated by a potential difference V_0 . Finally, they are bent into a circular path, of radius R , by a magnetic field B . Measurement of V_0 , R , and B allows one to calculate the ions' mass, m , as we will see directly. Since m is the mass of the ion, the mass of the corresponding neutral atom is calculated by adding one electron mass, m_e ; that of the bare nucleus is found by subtracting $(Z - 1)m_e$.

To calculate the mass, we note that when the ions are accelerated by the potential difference V_0 , they gain a kinetic energy,

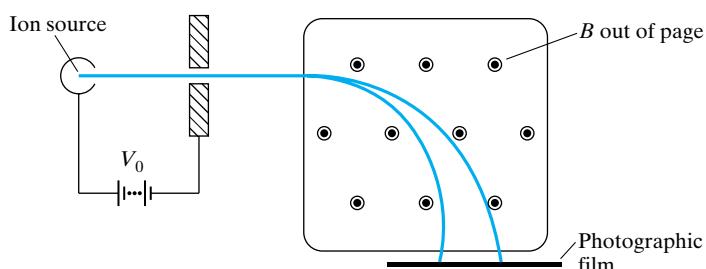
$$K = V_0 e$$

Their momentum p is then*

$$p = \sqrt{2mK} = \sqrt{2mV_0e} \quad (16.42)$$

FIGURE 16.18

A mass spectrometer. Positive ions are produced in the ion source by an electric discharge and are accelerated by a potential difference V_0 . They then enter a magnetic field and follow a curved path. Ions with different masses follow paths of different curvature and strike the photographic film at different places. The entire apparatus operates in a vacuum.



* We are treating the motion as nonrelativistic, which is usually an excellent approximation (see Problem 16.60).

Section 16.9 • Mass Spectrometers 29

where m is the mass of the ion. When the ions enter the magnetic field B , they follow a circular path with radius given by (2.48) as

$$R = \frac{P}{eB} \quad (16.43)$$

Substituting (16.42) and solving for m , we find that

$$m = \frac{eB^2 R^2}{2V_0} \quad (16.44)$$

If B and V_0 are fixed, different masses will have different radii, and measurement of R determines m from (16.44). In most early mass spectrometers, R was measured by locating the spots where the ions struck a photographic film, as in Fig. 16.18. In modern spectrometers a computer monitors the accelerating voltage V_0 , the field B , and the location of an electronic ion detector; after initial calibration with ions of known mass, the computer can display the results directly in atomic mass units.

When an element is analyzed in a mass spectrometer, its various isotopes are separated because of their different masses. Figure 16.19 shows the spectrum of masses obtained when natural lead is analyzed in a modern spectrometer, which displays its results directly as a graph of number of particles against mass.

Today, mass spectrometers are widely used in chemical analysis to help determine the composition of molecules. This is illustrated in Fig. 16.20, which shows the spectrum obtained when the simple diatomic compound HCl is introduced into the ion source of a mass spectrometer. Since chlorine has two stable isotopes, ^{35}Cl and ^{37}Cl , there are two large peaks with masses close to 36 u ($^1\text{H}^{35}\text{Cl}$) and 38 u ($^1\text{H}^{37}\text{Cl}$). In addition, smaller peaks can be seen at the positions one would expect for single atoms of ^{35}Cl and ^{37}Cl . These occur because the electric discharge in the ion source dissociates some of the molecules into H and Cl atoms. If the HCl analyzed in the spectrometer were an unknown substance, its composition could be deduced from the following observations: (1) the total molecular mass has two values, 36 and 38; (2) the molecule has fragments with masses of 35 and 37; and (3) the ratio of the peak heights at 36 and 38 (and those at 35 and 37) is the same as that of naturally occurring chlorine.

The mass spectrum obtained for a more complex molecule is shown in Fig. 16.21, together with a diagram of the molecule's structure. The molecule shown, methionine, has mass 149 u and produces the peak labeled 149 in the spectrum. As was the case with HCl, the molecule can be broken into smaller fragments in the ion source, and it is these fragments that produce the several peaks at lower masses. For example, the peaks labeled 61 and 88 arise because the molecule breaks easily at the point indicated by the arrow in Fig. 16.21(b), producing fragments of masses 61 u and 88 u. The masses of the various segments of this complex organic molecule provide valuable clues to its structure.

Close inspection of the major peaks in Fig. 16.21 shows that each is accompanied by several smaller peaks. These smaller peaks differ from the dominant peaks by one or two mass units and come about in two ways: (1) a molecule or molecular fragment can gain or lose one or two hydrogen atoms in the ion source; and (2) about 1% of the carbon atoms in any natural sample is the isotope ^{13}C (as opposed to the dominant ^{12}C). The details of these subsidiary peaks are additional clues to the composition of the molecule.

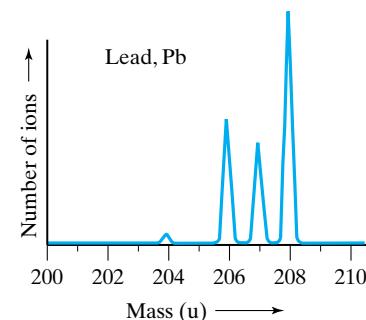


FIGURE 16.19

Results of analyzing natural lead in a mass spectrometer. The vertical axis shows the number of ions detected, and the horizontal axis is calibrated in atomic mass units. The different heights of the four peaks reflect the different natural abundances of the four isotopes of lead.

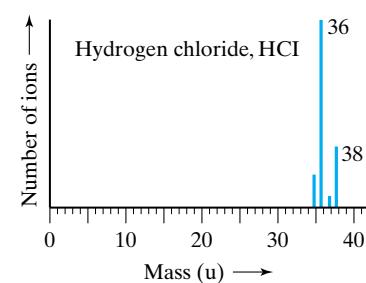
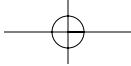


FIGURE 16.20

Mass spectrum obtained from HCl. The peaks labeled 36 and 38 were produced by the molecules $^1\text{H}^{35}\text{Cl}$ and $^1\text{H}^{37}\text{Cl}$, containing the two naturally occurring isotopes ^{35}Cl and ^{37}Cl . The smaller peaks at 35 and 37 were produced by atoms of ^{35}Cl and ^{37}Cl that were detached in the ion source.



30 Chapter 16 • The Structure of Atomic Nuclei

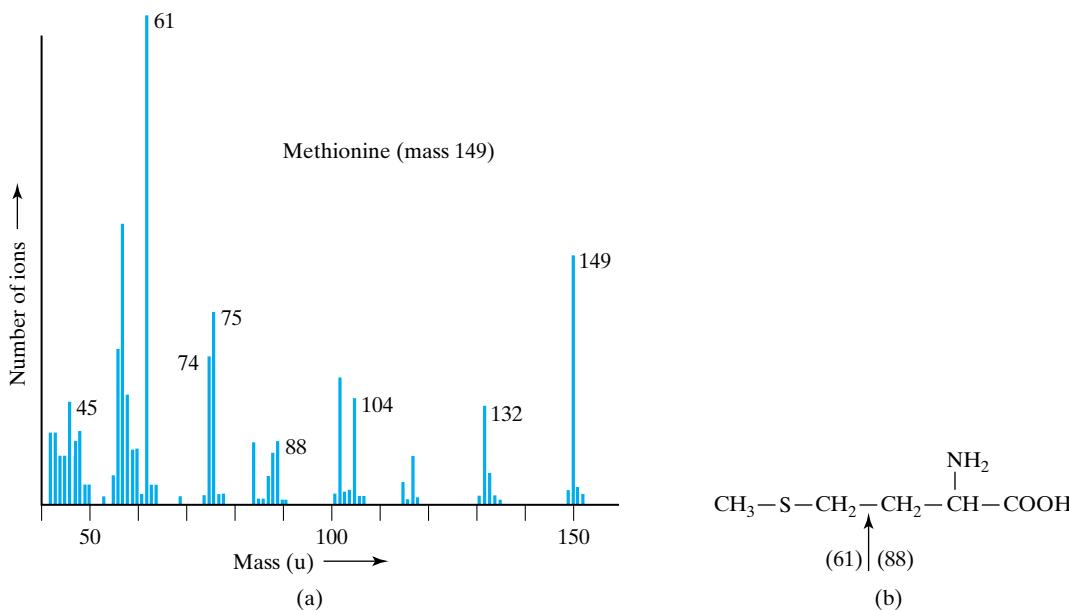


FIGURE 16.21

(a) Mass spectrum obtained from methionine, $C_5H_{11}O_2NS$, with mass 149 u. The lower mass peaks are produced by fragments of the molecule that were detached in the ion source. (b) The molecule's structure. The arrow shows a point where the molecule can easily be broken in the ion source; the resulting fragments have masses of 61 u and 88 u and produce the peaks labeled 61 and 88 in the spectrum.

Current research in molecular structure has extended mass spectroscopy to molecules with masses of thousands of atomic mass units. Mass spectrometers with this capability are commercially available and can quickly and automatically produce spectra like those in Figs. 16.20 and 16.21 for almost any substance that needs to be analyzed.

CHECKLIST FOR CHAPTER 16

CONCEPT	DETAILS
Isotopes	Different nuclei (or atoms) with the same atomic number Z
Isobars	Different nuclei (or atoms) with the same mass number A
Isotones	Different nuclei (or atoms) with the same neutron number N
Nuclear radius	$R = R_0 A^{1/3}$ (16.3)
Nuclear (or strong) force	Range and strength; charge independence (Sec. 16.3)
IPA for nucleons	Average potential energy seen by an individual nucleon in a nucleus (Sec. 16.5)
Symmetry effect	Pauli principle \Rightarrow tendency for nuclei to have $Z \approx N$ (Sec. 16.6)
Semiempirical binding-energy formula	Eq. (16.30)
Binding energy per nucleon, B/A	Measure of nuclear stability, Fig. 16.14
Shell model*	Magic numbers, spin-orbit splitting, total angular momentum (Sec. 16.8)
Mass spectrometers*	Measure atomic masses by observing curvature in magnetic field (Sec. 16.9).

PROBLEMS FOR CHAPTER 16

SECTION 16.2 (Nuclear Properties)

- 16.1** • Use Appendix D to find the numbers of protons and neutrons in the following nuclei: ^1H , ^4He , ^7Li , ^{20}Ne , ^{40}Ar , ^{63}Cu , ^{206}Pb .
- 16.2** • Use Appendix D to find the numbers of protons and neutrons in the following nuclei: ^2H , ^4He , ^9Be , ^{29}P , ^{64}Zn , ^{209}Bi , ^{232}Th .
- 16.3** • Use Appendix D to find the chemical symbols and the natural percent abundances of the isotopes with **(a)** $Z = 2$, $A = 3$; **(b)** $Z = 6$, $A = 13$; **(c)** $Z = 26$, $A = 56$; **(d)** $Z = 82$, $A = 206$.
- 16.4** • Use Appendix D to find the chemical symbols and the half-lives of the four radioactive nuclei with **(a)** $Z = 1$, $A = 3$; **(b)** $Z = 6$, $A = 14$; **(c)** $Z = 19$, $A = 40$; **(d)** $Z = 94$, $A = 244$.
- 16.5** • **(a)** Use Appendix D to make a list of all the stable isotopes, isobars, and isotones of ^{70}Ge . **(b)** Do the same for ^{27}Al and **(c)** ^{88}Sr .
- 16.6** • Make lists of all the stable isotopes, isobars, and isotones of each of the following: ^{56}Fe , ^{36}S , ^{208}Pb .
- 16.7** • Find the mass of the ^{12}C atom in Appendix D. What fraction of its mass is contained in its atomic electrons?
- 16.8** • The mass of a lead atom's nucleus is 3.5×10^{-25} kg, and its half-density radius R is 6.5 fm. Estimate its density by assuming its volume to be $(4/3)\pi R^3$. Compare your result to the density of solid lead (11 g/cm³).
- 16.9** • Atomic radii range from a little less than 0.1 nm to 0.3 nm, while nuclear radii range from about 2 fm to about 7 fm. Taking the representative values $R_{\text{atom}} \approx 0.2$ nm and $R_{\text{nuc}} \approx 4.5$ fm, find the ratio of the volume of an atom to that of a nucleus. What is the ratio of the corresponding average densities?
- 16.10** • Use the formula (16.3), $R = R_0 A^{1/3}$, with $R_0 = 1.07$ fm, to find the approximate radii of the most abundant isotopes of carbon, iron, and lead. (Mass numbers and abundances are in Appendix D.)
- 16.11** • The nuclear radius R is well approximated by the formula (16.3), $R = R_0 A^{1/3}$, where $R_0 = 1.07$ fm. Find the approximate radii of the most abundant isotopes of each of the following elements: Al, Cu, and U. (Mass numbers and abundances are given in Appendix D.)
- 16.12** • The first excitation energy of the ^7Li nucleus is 0.48 MeV. What is the wavelength of a photon emitted when a ^7Li nucleus drops from its first excited state to the ground state? Find the wavelength for the corresponding transition in the Li *atom* (excitation energy, 1.8 eV).
- 16.13** • The first excitation energy of the helium nucleus, ^4He , is about 20 MeV. What is the wavelength of a photon that can just excite the helium nucleus from its ground state? Compare this with the wavelength

of a photon that can just excite a helium *atom* (excitation energy 19.8 eV).

- 16.14** •• **(a)** Treating the nucleus as a uniform sphere of radius $(1.07 \text{ fm})A^{1/3}$, calculate the density (in kg/m³) of the ^{56}Fe nucleus and compare it to that of solid iron (7800 kg/m³). **(b)** Calculate the electric charge density of this nucleus ($Z = 26$) in coulombs/m³. Compare this charge density to that which can be uniformly distributed in an insulating sphere of 1 m radius in air without sparking. The maximum possible electric field strength at the surface of the sphere without sparking is about 3×10^6 V/m.

- 16.15** •• The density of mass inside a nucleus is shown in Fig. 16.2. There is no simple theory that predicts the exact shape shown, but it is found that the shape can be approximated by the following mathematical form, known as the Fermi function:

$$\rho(r) = \frac{\rho_0}{1 + e^{(r-R)/t}} \quad (16.45)$$

where ρ_0 , R , and t are positive constants, with $t \ll R$. **(a)** Prove that $\rho = \rho_0/2$ at $r = R$. **(b)** Show that the maximum value of ρ occurs at $r = 0$ and is very close to ρ_0 , given that $t \ll R$. **(c)** Sketch ρ as a function of r . **(d)** Prove that as r increases, the density falls from 90% to 10% of ρ_0 in a distance $\Delta r = 4.4t$. (Thus t characterizes the *thickness* of the surface region.)

SECTION 16.3 (The Nuclear Force)

- 16.16** • If there were no nuclear attraction, what would be the acceleration of either of two protons released at a separation of 4 fm? Compare your answer with g , the acceleration of gravity. (As large as your answer is, it is still small compared to nuclear accelerations. See Problem 16.20.)

- 16.17** • Use Eq. (16.7) to estimate the minimum kinetic energy of a nucleon in a nucleus of diameter 8 fm.

- 16.18** •• Use Eq. (16.7), taking the size a to be twice the radius given by (16.3), to find the minimum kinetic energy of a nucleon inside each of the nuclei ^{12}C and ^{244}Pu .

- 16.19** •• **(a)** The minimum kinetic energy K_{\min} of a nucleon in a rigid cubical box of side a is given by (16.7). By setting $a \approx 2R$, the nuclear diameter, and $R = R_0 A^{1/3}$ (with $R_0 = 1.07$ fm), derive a formula for K_{\min} as a function of the mass number A . **(b)** Find K_{\min} for ^9Be , ^{27}Al , and ^{238}U .

- 16.20** •• A representative value for the kinetic energy of a nucleon in a nucleus is 20 MeV. To illustrate the great strength of the nuclear force, do the following classical calculation: Suppose that a nucleon in a nucleus oscillates in simple harmonic motion with amplitude about 4 fm and peak kinetic energy 20 MeV. **(a)** Find its maximum acceleration. **(b)** By how many orders of magnitude does this exceed the acceleration of gravity, g ? **(c)** Find the force required to produce this acceleration.

32 Chapter 16 • The Structure of Atomic Nuclei

16.21 •• The charge independence of nuclear forces implies that in the absence of electrostatic forces, the energy levels of ^7Li and ^7Be would be the same. The main effect of the electrostatic forces is simply to raise all the levels of ^7Be compared to those of ^7Li . Approximating both nuclei as uniform spheres of charge $Q = Ze$ and the same radius R , estimate the difference in the binding energies for any level of ^7Be and the corresponding level of ^7Li . (The electrostatic energy of a uniform charged sphere is $3kQ^2/5R$ — see Problem 16.45. The observed radius R is about 2.5 fm.) Compare your rough estimate with the observed difference of about 1.7 MeV. (Note that the true charge distribution is not a uniform sphere with a well-defined radius R , but is spread out somewhat beyond R . Therefore, the observed electrostatic energies would be expected to be somewhat smaller than your estimates.)

16.22 •• Do the same calculations as in Problem 16.21, but for the nuclei ^{23}Na and ^{23}Mg . The observed difference is 4.8 MeV; use the formula $R = R_0 A^{1/3}$ to get the nuclear radius.

SECTION 16.4 (Electrons Versus Neutrons as Nuclear Constituents)

16.23 • One argument against the proton-electron model of the nucleus concerns the total spins of nuclei. The proton, electron, and neutron all have spin $\frac{1}{2}$, and the total spin of any number of spin-half particles takes the familiar form $\sqrt{s(s+1)}\hbar$. If there is just one particle, then of course, $s = \frac{1}{2}$. With two particles, the spins can be parallel or antiparallel, giving $s = 1$ or 0. It can be shown that the general rule is this: For an *odd* number of particles, the total spin has some half-odd-integer value for $s(s = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots)$. For an *even* number of particles, the total spin has s equal to some integer ($s = 0, 1, 2, \dots$). In light of this rule, consider the nucleus ^{14}N . **(a)** Assuming that ^{14}N is made of 7 protons and 7 neutrons, predict the character of its total spin (integer or half-odd integer). **(b)** Repeat for the proton-electron model of ^{14}N . **(c)** The observed total spin of ^{14}N is integer; which model does this support?

16.24 • Repeat Problem 16.23 for the deuteron ^2H , whose observed total spin is given by $s = 1$.

16.25 •• Supposing that nuclei could contain electrons, estimate the potential energy of one such electron at the center of a nucleus of ^{27}Al . Compare your answer with the minimum kinetic energy (16.10) of an electron in a nucleus. [Hint: The required potential energy is $-eV(0)$, where $V(r)$ is given by Eq. (16.49) below. The radius R is given by (16.3), and the total charge acting on the electron is $Q = (Z + 1)e$.]

16.26 •• (a) Following the suggestions in Problem 16.25, find the potential energy that an electron would have at the center of a $^{12}_6\text{C}$ nucleus. **(b)** Repeat for $^{208}_{82}\text{Pb}$. **(c)** Compare your answers with the minimum kinetic energy, which is of order 100 MeV.

16.27 •• Because of the proton's magnetic moment, every energy level of the hydrogen atom, as calculated previously, actually consists of two levels, very close together. The proton's magnetic moment creates a magnetic field, which means that the energy of the H atom is slightly different, depending on the relative orientations of the electron and proton moments — an effect known as **hyperfine splitting**. You can estimate the magnitude of the hyperfine splitting as follows: **(a)** The magnetic field at a distance r from a magnetic moment μ is

$$\mathbf{B} = \frac{\mu_0 \mu}{2\pi r^3} \quad (16.46)$$

where $\mu_0 = 4\pi \times 10^{-7} \text{ N/A}^2$ is the permeability of space. [For simplicity, we have given the field at a point on the axis of μ . At points off the axis the field is somewhat different, but is close enough to (16.46) for the purposes of this estimate.] Given that the magnetic moment of the proton is roughly equal to the nuclear magneton μ_N defined in (16.12), estimate the magnetic field at a distance a_B from a proton. **(b)** Taking your answer in part (a) as an estimate for the B field “seen” by an electron in the $1s$ state of a hydrogen atom, show that the atom's energy differs by roughly 10^{-6} eV for the cases that the proton and electron spins are parallel or antiparallel. [The energy of a magnetic moment in a B field is given by (9.10).] **(c)** This means that the $1s$ state of hydrogen is really two very closely spaced levels. Compare your rough estimate with the actual separation of these levels given that the photon emitted when a hydrogen atom makes a transition between them has $\lambda = 21 \text{ cm}$. (This 21-cm radiation is used by astronomers to identify hydrogen atoms in interstellar space.)

16.28 ••• It is often a useful approximation to treat a nucleus as a uniformly charged sphere of radius R and charge Q . In this problem you will find the electrostatic potential $V(r)$ inside such a sphere, centered at the origin. **(a)** Write down the electric field $E(r)$ at any point a distance r from the origin with $r > R$. (Remember that this is the same as the field of a point charge Q at the origin.) **(b)** Use the definition

$$\Delta V = V(r_2) - V(r_1) = - \int_{r_1}^{r_2} E(r) dr \quad (16.47)$$

to find the potential difference between points at r_1 and r_2 (both greater than R). It is usual to define $V(r)$ so that $V(r)$ approaches 0 as $r \rightarrow \infty$. By choosing $r_1 = \infty$ and $r_2 = r$, show that

$$V(r) = \frac{kQ}{r} \quad (r \geq R) \quad (16.48)$$

(c) Now find $E(r)$ for $r \leq R$. (Remember that tells us Gauss's law this is kQ'/r^2 , where Q' is the total charge *inside* the radius r .) Check that your answers for parts (c) and (a) agree when $r = R$. **(d)** Use Eq. (16.47) to find $V(r_2) - V(r_1)$ for any two points inside the sphere. Now choose $r_2 = r$ and $r_1 = R$, and use the value of $V(R)$ from part (b) to prove that

$$V(r) = \frac{kQ}{2R} \left(3 - \frac{r^2}{R^2} \right) \quad (r \leq R) \quad (16.49)$$

In Problems 16.25, 16.26, 16.29, and 16.30 this result is needed to find the potential energy of a charge inside a nucleus.

SECTION 16.5 (The IPA Potential Energy for Nucleons)

16.29 • (a) Use the result (16.49) of Problem 16.28 to estimate the electrostatic potential energy of a proton at the center of a ^{12}C nucleus. **(b)** Repeat for a nucleus of ^{208}Pb .

16.30 • The electrostatic potential energy U_{Coul} of a proton in a nucleus can be approximated by assuming that it moves in a uniform sphere of charge $(Z-1)e$ and radius R . **(a)** Use the results (16.48) and (16.49) of Problem 16.28 to write down and sketch $U_{\text{Coul}}(r)$. **(b)** Prove that $U_{\text{Coul}}(0) = 1.5 U_{\text{Coul}}(R)$. **(c)** Setting $R = (1.07 \text{ fm})A^{1/3}$, estimate $U_{\text{Coul}}(0)$ for a proton in ^4He and for a proton in ^{238}U .

SECTION 16.6 (The Pauli Principle and the Symmetry Effect)

16.31 • Enlarge Fig. 16.11 to include the nuclei ^{12}Be and ^{12}O . By how much (in terms of the quantity ΔE shown in Fig. 16.11) does the energy of each isobar with $Z \neq N$ exceed the energy of ^{12}C ?

SECTION 16.7 (The Semiempirical Binding-Energy Formula)

16.32 • Use the data of Appendix D to find the binding energy of ^{12}C .

16.33 • (a) Find the mass (in u) of the ^4He atom in Appendix D. **(b)** Find the mass of the ^4He nucleus to seven figures (but ignore corrections due to the atomic electrons' binding energy). **(c)** Do any of these seven figures change if you take into account the electrons' binding energy (about 80 eV total)?

16.34 • Use the data of Appendix D to find the binding energies of ^4He , ^{40}Ca , and ^{120}Sn .

16.35 • Use the data of Appendix D to deduce the energy required to remove one neutron from ^{14}C to produce ^{13}C and a neutron at rest, well separated from the ^{13}C nucleus.

16.36 • (a) Use conservation of energy to write an equation relating the mass of a nucleus (Z, N) and its neutron separation energy $S_n(Z, N)$ to the masses of the nucleus ($Z, N-1$) and the neutron. **(b)** Use the data in Appendix D to find S_n for ^{13}C , ^{120}Sn , and ^{200}Hg .

16.37 • (a) The proton separation energy S_p (energy to remove one proton) for ^{198}Hg is 7.1 MeV. Given that the total binding energy of ^{197}Au is 1559.4 MeV, find the total binding energy of ^{198}Hg . **(b)** Compare your answer with the answer obtained directly from the mass of ^{198}Hg given in Appendix D.

16.38 •• (a) To estimate the importance of the surface term in the binding-energy formula, consider a model of ^{27}Al that consists of 27 identical cubical nucleons packed into a $3 \times 3 \times 3$ cube. Each small cube has six faces. What is the fraction of all these faces that are exposed on the nuclear surface? **(b)** Repeat for ^{125}Te considered as a $5 \times 5 \times 5$ cube. **(c)** Comment on the difference in your answers.

16.39 •• The Coulomb energy of the charges in a nucleus is positive and hence reduces the binding energy B . If we approximate the nucleus as a uniform sphere of charge $Q = Ze$, then according to Eq. (16.50) of Problem 16.45, the Coulomb correction to B should be

$$B_{\text{Coul}} = -U_{\text{Coul}} = -\frac{3}{5} \frac{k(Ze)^2}{R}$$

Putting $R = R_0 A^{1/3}$, show that this agrees with the form (16.26) of the Coulomb correction given in Section 16.7. Use your answer here to calculate a_{Coul} in MeV, and compare with the empirical value $a_{\text{Coul}} = 0.711 \text{ MeV}$. (You should not expect precise agreement since the nucleus is certainly not a perfectly uniform sphere, but you should get the right order of magnitude.)

16.40 •• (a) Use the binding-energy formula (16.30) to predict the binding energy of ^{40}Ca . **(b)** Compare your answer with the actual binding energy found from the masses listed in Appendix D.

16.41 •• Use the semiempirical binding-energy formula to find B/A for ^{20}Ne , ^{60}Ni , ^{259}No . Make a table showing all five terms for each nucleus, and comment on their relative importance. Do your final answers follow the general trend of Fig. 16.14?

16.42 •• Use the binding-energy formula (16.30) to find the total binding energies of ^{56}Ni , ^{56}Co , ^{56}Fe and ^{56}Mn . Which terms are different in the four cases, and why?

16.43 •• If ^{235}U captures a neutron to form ^{236}U in its ground state, the energy released is $B(^{236}\text{U}) - B(^{235}\text{U})$. **(a)** Prove this statement. **(b)** Use the binding-energy formula (16.30) to estimate the energy released, and compare with the observed value of 6.5 MeV. (Note: We have assumed here that ^{236}U is formed in its ground state, and the 6.5 MeV is carried away, by a photon, for example. An important alternative is that ^{236}U can be formed in an excited state, 6.5 MeV above the ground state. This excitation energy can lead to oscillations that cause the nucleus to fission.)

16.44 •• Compare the surface area of a sphere to that of a cube of the same volume. How much energy, in MeV, would be needed to distort the normally spherical ^{60}Ni nucleus into a cubical shape if the volume and Coulomb energies changed by negligible amounts?

16.45 ••• In this problem you will calculate the electrostatic energy of a uniform sphere of charge Q with radius R . The electric potential $V(r)$ at any radius $r < R$ is given by (16.49) in Problem 16.28, and the potential energy of a charge q at radius r is $qV(r)$. **(a)** Write down the total charge contained between radius r and

34 Chapter 16 • The Structure of Atomic Nuclei

$r + dr$, and hence find the potential energy of that charge. (b) If you integrate your answer to (a) from $r = 0$ to $r = R$, you will get twice the total potential energy of the whole sphere since you will have counted the energy of any two charge elements twice. Show that the total Coulomb energy of the sphere is

$$U_{\text{Coul}} = \frac{3}{5} \frac{kQ^2}{R} \quad (16.50)$$

- 16.46 •••** (a) Substitute the binding-energy formula (16.30) into the relation

$$m_{\text{nuc}} = Zm_p + Nm_n - \frac{B}{c^2}$$

to obtain the semiempirical *mass* formula. [To simplify matters, take A to be odd, so that the pairing term in (16.30) is zero.] (b) Among any set of isobars, the nucleus with lowest mass is the most stable. To identify this nucleus, first write your mass formula in terms of the variables A and Z (that is, replace N , wherever it appears, by $A - Z$). With A fixed, differentiate m with respect to Z . The minimum mass is determined by the condition $\partial m / \partial Z = 0$. Show that this leads to a relation of the form

$$Z = \frac{A}{2} \frac{1 + \alpha}{1 + \beta A^{2/3}} \quad (16.51)$$

where α and β are related to the coefficients (16.31) as follows:

$$\alpha = \frac{(m_n - m_p)c^2}{4a_{\text{sym}}} \quad \text{and} \quad \beta = \frac{a_{\text{Coul}}}{4a_{\text{sym}}}$$

- (c) For $A = 37, 115, 185$, find the values of Z that give the most stable nuclei and compare with the observed values from Appendix D.

SECTION 16.8 (The Shell Model*)

- 16.47 •** Which of the following nuclei has a closed subshell for Z ? Which for N ? ${}^{10}\text{B}$, ${}^{11}\text{B}$, ${}^{17}\text{O}$, ${}^{27}\text{Al}$, ${}^{28}\text{Si}$, ${}^{42}\text{Ca}$, ${}^{48}\text{Ca}$.

- 16.48 •** Use the levels shown in Fig. 16.17 to predict j_{tot} for the ground states of the nine nuclei ${}^{40}\text{Ca}$, ${}^{41}\text{Ca}$, ..., ${}^{48}\text{Ca}$. Compare with the data in Appendix D.

- 16.49 •** Use the levels shown in Fig. 16.17 to predict the total angular momenta of the following nuclei: ${}^{29}\text{Si}$, ${}^{33}\text{S}$, ${}^{37}\text{Cl}$, ${}^{59}\text{Co}$, and ${}^{71}\text{Ga}$.

- 16.50 •** Use the shell model to predict j_{tot} for the ground states of ${}^{39}\text{K}_{20}$, ${}^{40}\text{Ca}_{20}$, and ${}^{41}\text{Sc}_{20}$.

- 16.51 •** The ground state of ${}^7\text{Li}$ has $j_{\text{tot}} = \frac{3}{2}$ and its first excited state has $j = \frac{1}{2}$. Explain these observations in terms of the shell model.

- 16.52 • (a)** Use the levels shown in Fig. 16.17 to draw an occupancy diagram like Fig. 16.10(b) for the ground state of ${}^7\text{Be}$. (b) What is j_{tot} ? (c) What would you predict for the first excited state? (Draw an occupancy diagram, and give j_{tot} .)

- 16.53 •** In applying the shell model, we usually use the level ordering of Fig. 16.17, which is based on an assumed spherical potential well. For certain nonspherical nuclei, this assumption is incorrect and the ordering of the levels is different. Examples are ${}^{19}\text{F}$ and ${}^{121}\text{Sb}$; check this by using the rule (16.41) and Fig. 16.17 to predict the angular momenta of these two nuclei and comparing with the observed values in Appendix D.

- 16.54 •• (a)** Use the energy levels of Fig. 16.17 to draw an occupancy diagram like Fig. 16.10(b) for ${}^{13}\text{C}$. (b) What is j_{tot} for the ground state? (c) The first excited state is formed by lifting one neutron from $1p_{3/2}$ to $1p_{1/2}$. Draw an occupancy diagram for this state, and explain what you would predict for j_{tot} of this state. (Does your answer agree with the observed value, $j_{\text{tot}} = \frac{3}{2}$?) (d) Now predict j_{tot} for the ground and first excited states of ${}^{13}\text{N}$. (In this case the excitation involves a proton.)

- 16.55 ••** For each choice of n and l (with $l \neq 0$), a nucleon has two different possible energy levels with $j = l \pm \frac{1}{2}$. Prove that the sum of the degeneracies of these two levels is equal to the total degeneracy that the level nl would have had in the absence of any $\mathbf{S} \cdot \mathbf{L}$ splitting.

- 16.56 •••** The first magic number for both nuclei and atoms is 2. The second is 8 for nuclei and 10 for atoms. The difference occurs because what we call the $1p$ and $2s$ levels in nuclei are widely separated, whereas the corresponding levels in atoms (called $2s$ and $2p$ by atomic physicists) are very close together. To explain this difference, compare the shape of an atomic potential energy ($\propto 1/r$) with that of the nuclear potential well. Noting that states with higher l have probability distributions that are pushed out to larger radii, qualitatively explain the difference in the positions of these two levels.

- 16.57 •••** The spin-orbit energy (16.34) is proportional to $\mathbf{S} \cdot \mathbf{L}$. To evaluate $\mathbf{S} \cdot \mathbf{L}$, note that since $\mathbf{J} = \mathbf{S} + \mathbf{L}$,

$$J^2 = (\mathbf{S} + \mathbf{L}) \cdot (\mathbf{S} + \mathbf{L}) = S^2 + L^2 + 2\mathbf{S} \cdot \mathbf{L}$$

You can solve this for $\mathbf{S} \cdot \mathbf{L}$ and then put in the values for J^2 , S^2 , and L^2 . [For instance, $J^2 = j(j+1)\hbar^2$ where $j = l \pm \frac{1}{2}$.] Show that if $j = l + \frac{1}{2}$ then $\mathbf{S} \cdot \mathbf{L} = l\hbar^2/2$, and if $j = l - \frac{1}{2}$, then $\mathbf{S} \cdot \mathbf{L} = -(l+1)\hbar^2/2$. Use these answers to prove that the spin-orbit splitting increases with increasing l , as indicated in Fig. 16.17.

SECTION 16.9 (Mass Spectrometers*)

- 16.58 •** Where does the methionine molecule in Fig. 16.21 break to account for the two peaks at 74 u and 75 u in its mass spectrum?

- 16.59 •** Repeat Problem 16.58 for the peaks at 45 u and 104 u.

- 16.60 •• (a)** Find the radius R of the curved path for a once-ionized nitrogen atom in a mass spectrometer with $B = 0.05$ T and $V_0 = 10$ kV, using (16.44).

(b) Equation (16.44) was derived nonrelativistically. What would R be if calculated relativistically? How many significant figures must you include to see the difference?

- 16.61 ••** Conditions in the ion source of a mass spectrometer can be adjusted to allow formation of multiply ionized atoms. **(a)** Rewrite Eq. (16.44) for the case of an n times ionized atom, and solve for R . **(b)** Taking the masses of ${}^4\text{He}$, ${}^{12}\text{C}$, and ${}^{16}\text{O}$ to be 4 u, 12 u, and 16 u, respectively, write down expressions for the radii, R , for He^+ , C^{3+} , and O^{4+} . **(c)** Using the actual masses of the atoms, find the fractional differences in these three radii (He versus C and C versus O). The tiny differences in these radii can be measured very accurately and allow precise comparisons of the three masses involved (one of which, ${}^{12}\text{C}$, defines the atomic mass scale).

- 16.62 ••** The chemical composition of a molecule can often be deduced from a very accurate measurement of its mass. As a simple example consider the following. Gas from a car's exhaust is analyzed in a mass spectrometer. Two of the resulting peaks are very close to 28 u. **(a)** Can you suggest what gases produced these two peaks? **(b)** Careful measurements indicate that the masses of the neutral molecules concerned are 28.006 u and 27.995 u. What are the two molecules?

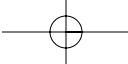
COMPUTER PROBLEMS

16.63 •• (Section 16.2) Same as Problem 16.15, but in part (c) use graph-plotting software to plot $\rho(r)$ for ${}^{208}\text{Pb}$ from $r = 0$ to $r = 15$ fm, given the following parameters: $\rho_0 = 3.25 \times 10^{17} \text{ kg/m}^3$, $t = 0.55$ fm, and $R = R_0 A^{1/3}$, where $R_0 = 1.07$ fm. Plot $\rho(r)$ in units of ρ_0 ; that is, plot $\rho(r)/\rho_0$.

16.64 •• (Section 16.2) Same as Problem 16.63, but make plots for ${}^{12}\text{C}$ and ${}^{244}\text{Pu}$, drawing both curves on the same plot.

16.65 •• (Section 16.2) Same as Problem 16.63, but use suitable plotting software to make plots for ${}^4\text{He}$ and ${}^{238}\text{U}$.

16.66 ••• (Section 16.7) **(a)** According to the semiempirical binding-energy formula, the most stable value of Z for any fixed A is given by (16.51) in Problem 16.46. (This is for A odd.) Using suitable plotting software, plot the curve of stability, showing the most stable values of (N, Z) for $0 < A < 240$, as predicted by (16.51). The simplest way to do this is to make a parametric plot of (N, Z) as a function of the parameter A . Include on the same plot the positions of the following odd- A nuclei: ${}^7\text{Li}$, ${}^{63}\text{Cu}$, ${}^{107}\text{Ag}$, ${}^{187}\text{Au}$, and ${}^{235}\text{U}$.



Chapter 17

Radioactivity and Nuclear Reactions

- 17.1** Introduction
- 17.2** Radioactivity
- 17.3** The Exponential Decay Law
- 17.4** Beta Decay and the Neutrino
- 17.5** The Natural Radioactive Series
- 17.6** Nuclear Reactions
- 17.7** Nuclear Fission
- 17.8** Nuclear Fusion
- 17.9** Nuclear Weapons
- 17.10** The Theory of Alpha Decay*
- 17.11** Accelerators*
- 17.12** Particle Detectors*
- 17.13** Units of Radiation*

Problems for Chapter 17

*These sections can be omitted without serious loss of continuity.

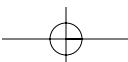
17.1 Introduction

In Chapter 16 we discussed the structure of isolated, static nuclei — their constituents, masses, energy levels, and angular momenta. In this chapter we discuss various nuclear processes, in which nuclei change their state — and often their identity as well — by emitting and absorbing energy in the form of photons or other particles. Historically, the first of these processes to be studied was the decay of naturally occurring unstable nuclei. Such nuclei were called *radioactive*, and we discuss radioactivity in Sections 17.2 through 17.5.

In Sections 17.6 through 17.9 we discuss nuclear reactions, the processes in which two nuclei collide and induce a variety of nuclear transformations. The two best-known examples of nuclear reactions, fission and fusion, are of great practical importance as sources of enormous energy, both peaceful and destructive, as we discuss in Sections 17.7 through 17.9.

In Section 17.10 we describe the quantum theory of the α decay of certain radioactive nuclei. This was an early application of the Schrödinger equation to nuclear physics. It is an example of the process called quantum tunneling that we discussed in Section 7.10.

In the Sections 17.11 and 17.12 we describe some of the devices used to accelerate and detect nuclear particles. In studying microscopic bodies such as small organisms, one shines light on them and examines the deflected light



with a microscope. Nuclei are, of course, far too small to be seen with ordinary light. Instead, nuclei are probed by bombarding them with energetic particles (electrons, protons, alpha particles, etc.) and observing the deflected particles with suitable detectors. The accelerators that boost the particles to the required energies and the devices that detect them can be regarded as the “light” sources and “microscopes” of the nuclear physicist. Finally, in Section 17.13 we describe the various units used to measure amounts of radiation.

If your only interest in this chapter is to prepare yourself for Chapter 18 on elementary particles, you need read only Sections 17.2 through 17.4 (and perhaps 17.11 and 17.12).

17.2 Radioactivity

In 1896 the French physicist Henri Becquerel discovered that uranium salts emit radiation that could darken a photographic film, even if the film was wrapped in black lightproof paper. Evidently, the radiation, just like ordinary light, could ionize the atoms in photographic emulsion and hence “expose” the film. Unlike ordinary light, however, it could pass through black paper. The emission of this new radiation came to be called **radioactivity**, and within two years Marie and Pierre Curie had identified three more radioactive elements: thorium, polonium, and radium. (The Curies and Becquerel shared the 1903 Nobel Prize for physics for their work on radioactivity.) We now know that radioactivity originates in the atomic nucleus. Thus, although the existence of nuclei was not established for another 15 years, we can date the beginning of nuclear physics from Becquerel’s discovery in 1896.

The radiation from radioactive materials was found to ionize air, and the measurement of electric currents passing through the ionized air became a common method of detecting and measuring the radiation. It was soon found that the radiation could be divided into three types, which Ernest Rutherford called α -, β -, and γ -rays. Alpha-rays ionize air the most and are the most easily stopped as they pass through matter. (They can be stopped by a single sheet of paper, for example.) Beta-rays are intermediate in the ionization they cause and in their ability to penetrate matter; gamma-rays ionize the least and penetrate farthest. (Beta-rays are stopped by a few millimeters of metal; gamma-rays penetrate an order of magnitude farther.*¹) Electric and magnetic deflections show that α -rays carry a positive charge and β -rays a negative charge, while γ -rays are neutral.

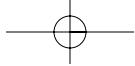
Using electric and magnetic deflections, much as Thomson had with electrons (Section 3.10), experimenters measured the mass-to-charge ratio, m/q , for α - and β -rays. For α -rays, it was found that $m/q = 2m_H/e$. This ratio could be explained if α -rays consisted of once-ionized hydrogen molecules, H_2^+ , or twice-ionized helium atoms, He^{++} , or other possibilities, such as six-times-ionized carbon 12. In a beautifully direct experiment, Rutherford established that He^{++} was the correct choice. He had a glass vessel fitted with two electrodes and part of its wall made thin enough that α -rays could enter its evacuated interior. When a source of α -rays was placed nearby for several days, a

Marie Curie
(1867–1934, Polish-French)



Marie Curie was the first person to win two Nobel Prizes. She shared the 1903 prize in physics with her husband, Pierre, and Henri Becquerel for work on radioactivity, and then won the 1911 prize in chemistry for the discovery of the elements polonium and radium. Her daughter and son-in-law, the Joliot-Curies, also won the Nobel Prize in chemistry (in 1935 for discovering man-made radioactivity).

*These differences are useful in identifying the different radiations. They originate in the different masses, speeds, and charges of the particles involved. The details are quite complicated and are not worth discussing here.



38 Chapter 17 • Radioactivity and Nuclear Reactions

detectable quantity of helium accumulated inside the vessel, as demonstrated by an electric discharge that produced the spectrum characteristic of helium. Thus alpha particles were shown to be He^{++} ions or, as we now know, He nuclei.

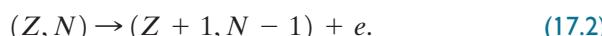
It was later established that the radiation emitted by radioactive materials originates in the atomic nucleus. All radioactive nuclei have excess energy, which makes them unstable. In α -radioactive nuclei the excess energy is released by the emission of an alpha particle. If the original nucleus has Z protons and N neutrons, we can write this process of **alpha decay** as



Notice that the remaining “offspring” nucleus has two fewer protons and two fewer neutrons than the original “parent” and is therefore the nucleus of a different element, with an atomic number two less than that of the parent. It was for establishing that radioactivity involved this transformation of elements that Rutherford won the Nobel Prize in chemistry in 1908.

In the case of β -rays it was found that $m/q = -m_e/e$, suggesting that β -rays were just electrons or “cathode rays,” as they were then called. In due course it was established that β -rays are indeed electrons, although the energy of β electrons is much larger than that of typical cathode rays (several MeV versus 100 keV at most). The energy of β particles is also much larger than their rest energy (0.5 MeV). Thus the electrons emitted by radioactive nuclei were the first particles observed moving at relativistic speeds (other than photons, of course).

The radioactive nuclei that emit β particles do so because they have too many neutrons; that is, the ratio, N/Z , of neutrons to protons is larger than the value that gives maximum stability. As mentioned in Chapter 16, one of the excess neutrons can change into a proton by creating and ejecting an electron. The resulting process of **beta decay** can be written as*



Like α decay, β decay causes a transformation of the element involved — in this case an increase of one in the atomic number Z .

The γ -rays emitted by radioactive nuclei were found to have many properties in common with X-rays, and it was eventually proved that, like X-rays, γ -rays are electromagnetic radiation, the only difference being their energies: X-ray photons (that is, photons produced in X-ray tubes) seldom exceed 100 keV or so, because of the difficulty of maintaining higher voltages; γ -ray photons have energies up to about 10 MeV, depending on the radioactive nucleus involved. As we mentioned in Chapter 16, the origin of γ -rays is exactly analogous to that of the photons given off by atoms: When a nucleus is in an excited state, it can emit a photon and drop to a lower level.

17.3 The Exponential Decay Law

The radiation from a radioactive sample consists of many individual particles, each of which came from its own separate parent nucleus. In this section we discuss the laws governing the rate at which these radioactive decays occur,

*As we discuss in Section 17.4, the equation (17.2) is missing one term on the right: In addition to the electron, a neutral particle called a neutrino is produced. This particle is so difficult to detect that it was not observed directly until 1956.

Section 17.3 • The Exponential Decay Law 39

and the variation of that rate as time goes by. Our discussion applies equally to all three kinds of radioactivity (α , β , and γ) and, in fact, to the decay of any unstable quantum system. Thus, the same ideas apply to the decay of unstable elementary particles (such as the pions discussed in Section 1.9) and to the decay of excited atomic states (as discussed in Chapter 11).

If the unstable nuclei of a given species were identical clock-like mechanisms obeying the laws of classical physics, we would expect all of them to decay at the same definite time after their formation. Instead, they are found to decay after a wide range of different times. The explanation of this behavior lies in the probabilistic nature of quantum mechanics. We have seen, for example, that quantum mechanics does not give the precise location of an electron in an atom; instead, it gives only the probabilities of the various possible locations. In the same way, quantum mechanics cannot predict the precise time at which a radioactive nucleus will decay; instead, it gives only the probabilities of the various times at which the decay may occur.

We will denote by r the probability that a particular nucleus will decay in any unit interval of time. This parameter is called the **decay constant** of the nucleus in question. If, for example, $r = 0.02$ per second (or 0.02 s^{-1}), the probability is 2% that the nucleus will decay in any 1-second interval. It is an experimental fact that for any given kind of nucleus, the probability r is a constant, independent of how long ago the nucleus was formed.* Given this fact, we can set up and solve the equation that determines the rate at which the nuclei in a sample will decay.

Suppose that at time t our sample contains a large number, $N = N(t)$, of radioactive nuclei, all of the same species (and hence with the same value of r). The decay constant r is the probability that any one nucleus will decay in unit time. Therefore, the total number of nuclei expected to decay in unit time is the probability r multiplied by the number of nuclei $N(t)$; that is, the total rate $R(t)$ at which nuclei will decay is

$$R(t) = rN(t) \quad (17.3)$$

Now, each time a nucleus decays, the number $N(t)$ decreases by 1. Thus, R is just the rate of decrease of N :

$$R = -\frac{dN}{dt} \quad (17.4)$$

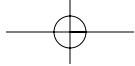
Combining (17.3) and (17.4), we find that

$$\frac{dN}{dt} = -rN \quad (17.5)$$

This is a first-order differential equation for $N(t)$ and asserts that the derivative of N is proportional to N itself. The only function that satisfies this equation is the exponential function (Problem 17.12)

$$N(t) = N_0 e^{-rt} \quad (17.6)$$

*We can re-express this informally by saying that a nucleus is unaware of its own past history: At the beginning of each 1-second interval it has the same probability of decay, whether it has been in existence for 1 millisecond or one century. There is a partial analogy here with the probabilities in throwing a die: Provided that the die is true, the probability of throwing a one is the same $1/6$, whether the last 10 throws were all ones or none of them were.



40 Chapter 17 • Radioactivity and Nuclear Reactions

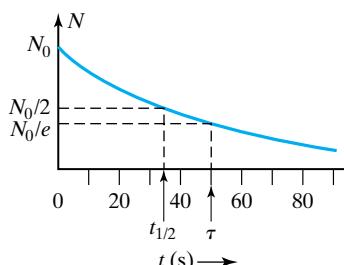


FIGURE 17.1

The number, N , of nuclei in a radioactive sample decreases exponentially, as in Eq. (17.6). At time $t = 0$, the number is $N = N_0$. At time $t_{1/2}$ it has dropped to $N_0/2$; at time τ , it has dropped to N_0/e . The decay constant is $r = 0.02 \text{ s}^{-1}$ in this example.

where N_0 is a constant. To identify N_0 , we set $t = 0$ and find that N_0 is simply the number of nuclei at time $t = 0$. Therefore, we have the following important conclusion: If a sample consists of N_0 radioactive nuclei at time $t = 0$ and if $N(t)$ denotes the number of the nuclei remaining at any later time t , then the function $N(t)$ decreases exponentially with time, as given by (17.6). This behavior is sketched for the case $r = 0.02 \text{ s}^{-1}$ in Fig. 17.1.

There are several ways to characterize the rate at which a radioactive nucleus decays. One is to give the decay constant, r , the probability that any one nucleus will decay in unit time. An alternative is to give the reciprocal, $1/r$, which is denoted by τ ,

$$\tau = \frac{1}{r} \quad (17.7)$$

Since r has the dimensions $(\text{time})^{-1}$, the parameter τ has the dimensions of time, and τ is called the **lifetime of the nucleus**. Its significance follows easily from (17.6): With $t = \tau = 1/r$,

$$N(\tau) = N_0 e^{-1} = \frac{N_0}{e}$$

that is, τ is *the time in which N drops to the fraction $1/e$ of its original value*, as indicated in Fig. 17.1. For this reason, τ is sometimes called the $1/e$ time of the nucleus. One can also show (Problem 17.13) that τ is the *average* time for which the nuclei in a sample survive, and τ is therefore also called the **mean life** of the nucleus. Note that in terms of τ , the decay law (17.6) can be rewritten

$$N(t) = N_0 e^{-t/\tau} \quad (17.8)$$

Another way to characterize the decay rate of a nucleus is to give its **half-life**, $t_{1/2}$, the time in which the number of nuclei drops to half its original value:

$$N(t_{1/2}) = \frac{1}{2} N_0 \quad (17.9)$$

According to (17.6), $t_{1/2}$ is determined by the equation

$$N_0 e^{-rt_{1/2}} = \frac{1}{2} N_0$$

Cancelling the N_0 and taking the natural logarithm of this equation, we find that

$$-rt_{1/2} = \ln\left(\frac{1}{2}\right) = -\ln 2$$

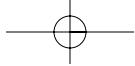
or

$$t_{1/2} = \frac{\ln 2}{r} = \frac{0.693}{r} \quad (17.10)$$

or since $\tau = 1/r$,

$$t_{1/2} = \tau \ln 2 = 0.693\tau \quad (17.11)$$

The half-life is also indicated on Fig. 17.1.

Section 17.3 • The Exponential Decay Law **41**

It is a simple exercise (Problem 17.11) to check that the exponential decay law (17.6) can be rewritten as

$$N(t) = \frac{N_0}{2^{t/t_{1/2}}} \quad (17.12)$$

In this expression the quantity $t/t_{1/2}$ is simply the number of half-lives that have gone by in time t :

$$n = (\text{number of half-lives in time } t) = \frac{t}{t_{1/2}} \quad (17.13)$$

Thus, (17.12) can be rewritten as

$$N(t) = \frac{N_0}{2^n} \quad (17.14)$$

that is, in n half-lives, the population halves itself n times over. This is the property of half-lives that was introduced in Section 1.9. Note that (17.12) can also be rewritten, in close parallel with (17.8), as

$$N(t) = N_0 2^{-t/t_{1/2}}$$

It is important to be aware that the parameters $t_{1/2}$ and τ are both in common use, and one needs to feel at home with either. In particular, nuclear physicists have traditionally preferred the half-life $t_{1/2}$, while particle physicists tend to use the mean life τ . Mathematically, the mean life is more convenient, but the popular press uses the term half-life exclusively.

Example 17.1

The nucleus of thorium 231 is β radioactive, with half-life $t_{1/2} = 25.6$ hours. At $t = 0$ a sample of ^{231}Th contains $N_0 = 56,000$ nuclei. How many ^{231}Th nuclei will remain after 3 days? What is the total rate at which electrons are emitted at $t = 0$? What will the rate be after 3 days?

To get a rough estimate for the number of nuclei after 3 days, we note that $t_{1/2}$ is about a day. Therefore, 3 days is about 3 half-lives, and N will have halved itself about three times:

$$N(3 \text{ days}) \approx \frac{N_0}{2^3} = \frac{56,000}{8} = 7000$$

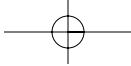
If we want a more accurate answer, we have only to calculate the number of half-lives more accurately:

$$n = \frac{t}{t_{1/2}} = \frac{72 \text{ hours}}{25.6 \text{ hours}} = 2.81$$

Therefore,

$$N(3 \text{ days}) = \frac{N_0}{2^{2.81}} = \frac{56,000}{7.01} = 8000 \quad (17.15)$$

to two significant figures.



42 Chapter 17 • Radioactivity and Nuclear Reactions

It is perhaps worth emphasizing that we could do the same calculation in terms of the exponential decay constant r , as given by (17.10),

$$r = \frac{\ln 2}{t_{1/2}} = 0.0271 \text{ hour}^{-1}$$

With $t = 3$ days = 72 hours, the exponential decay law (17.6) gives

$$N(3 \text{ days}) = N_0 e^{-rt} = (56,000) \times e^{-(0.0271 \times 72)} \approx 8000$$

as before.

To find the rate at which electrons are emitted, we note that each decay gives one electron, so that the rate of electron emission is the same as the rate R of decays. According to (17.3), this is

$$R(t) = rN(t) \quad (17.16)$$

which, with $t = 0$, gives

$$\begin{aligned} R_0 &= rN_0 = (0.0271/\text{hour}) \times 56,000 \\ &= 1520 \text{ electrons/hour} \approx 25 \text{ electrons/minute}. \end{aligned}$$

According to (17.16), the rate $R(t)$ is proportional to the number of nuclei, $N(t)$. Thus $R(t)$ shows the same exponential decay as $N(t)$, and in any time t the rate R decreases by the same factor as does N . In particular, we saw in (17.15) that in 3 days N drops to 1/7.01 of its original value; therefore, the same is true of R :

$$R(3 \text{ days}) = \frac{R_0}{7.01} = 3.6 \text{ electrons/minute}$$

Radioactive Dating

To conclude this section, we mention an application of radioactivity that makes explicit use of the exponential decay law: the dating of archeological and geological specimens. If a specimen contains a measurable amount of a known radioactive material (N atoms, say) and if one can reliably estimate the number, N_0 , of atoms that were present when the specimen was formed, then the ratio N/N_0 should be the exponential decay factor $\exp(-rt)$. Knowing N , N_0 , and r , one can calculate the age of the specimen.

Perhaps the best-known example is the method of radiocarbon dating. The isotope ^{14}C is β radioactive with a half-life of 5730 years. It occurs naturally because it is produced continually in collisions between cosmic rays and nitrogen nuclei in the atmosphere. The number of ^{14}C atoms in the atmosphere represents an equilibrium between the rate of production by cosmic rays and the rate of loss by decay. This number holds nearly constant throughout the atmosphere at about 1.3 atoms of ^{14}C in every 10^{12} atoms of all isotopes of carbon. Since ^{14}C is chemically indistinguishable from its stable isotopes, this proportion persists through all chemical reactions. Plants obtain their carbon from the atmosphere in the form of CO_2 and fix it by photosynthesis. Animals get their carbon when they eat plants or plant-eating animals. Thus, all living organisms contain ^{14}C in the same proportion as the atmosphere. However, as soon as an organism dies, its intake of fresh carbon ceases

Section 17.3 • The Exponential Decay Law 43

and the proportion of ^{14}C atoms starts to diminish exponentially with a known decay constant. By measuring the proportion of ^{14}C atoms in a specimen of wood, bone, or cloth, one can calculate how long it is since the specimen was part of a living organism.

Other radioactive nuclei are used similarly to date geological samples. In this way the oldest known rocks have been shown to have solidified about 4 billion years ago, and the oldest fossils formed about 3 billion years ago.

Example 17.2

In radiocarbon dating the amount of ^{14}C in a specimen is usually found by measuring the rate at which it ejects electrons (using a Geiger counter for example). If 30 grams of carbon from the wood from a prehistoric dwelling eject 200 electrons per minute, about how old is the dwelling?

When the 30 grams of carbon was part of a branch that had just been cut from a living tree, we know that the proportion of ^{14}C atoms was 1.3×10^{-12} . At that time (which we call $t = 0$) the number of ^{14}C atoms in the 30 grams of carbon was

$$\begin{aligned} N_0 &= (1.3 \times 10^{-12}) \times \frac{30 \text{ gram}}{12 \text{ gram/mole}} \times \left(6.02 \times 10^{23} \frac{\text{atoms}}{\text{mole}} \right) \\ &= 1.96 \times 10^{12} \text{ atoms} \end{aligned}$$

Since the half life of ^{14}C is $t_{1/2} = 5730$ years, the decay constant is

$$r = \frac{\ln 2}{t_{1/2}} = \frac{0.693}{5730 \times 365 \times 24 \times 60 \text{ min}} = 2.30 \times 10^{-10} \text{ min}^{-1}$$

Therefore, the total decay rate from our sample at $t = 0$ was

$$R_0 = rN_0 = (2.30 \times 10^{-10} \text{ min}^{-1}) \times (1.96 \times 10^{12}) = 451 \text{ decays/min}$$

Today (time t) the rate of decays is measured to be

$$R = 200 \text{ decays/min}$$

This rate is roughly half of R_0 . Thus the time that has elapsed since the wood was part of a living tree is about a half-life, or roughly 6000 years. More precisely, the number, n , of half-lives is found by solving the equation

$$\frac{R}{R_0} = \frac{1}{2^n}$$

Taking the logarithm of this equation, we find that

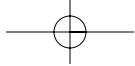
$$\log \frac{R_0}{R} = n \log 2$$

or

$$n = \frac{\log(R_0/R)}{\log 2} = \frac{\log(451/200)}{\log 2} = 1.17$$

That is, the age of the wood (and, hence, presumably the dwelling) is 1.17 half-lives, or

$$t = 1.17t_{1/2} = 1.17 \times (5730 \text{ years}) = 6700 \text{ years}$$



44 Chapter 17 • Radioactivity and Nuclear Reactions

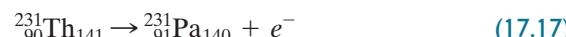
17.4 Beta Decay and the Neutrino

In several ways β decay is the most interesting of the radioactive decay processes. In particular, for 20 years there was a serious puzzle over the number of particles ejected in β decay. It is this puzzle and its resolution, with the discovery of the neutrino, that we describe in this section.

We saw in Section 16.6 that because of the Pauli principle, nuclei with $Z \approx N$ tend to have lower rest energy than their isobars with Z much different from N . This claim is easily verified using the experimentally measured masses of any series of isobars. For example, the rest masses of five isobars with $A = 11$ are plotted in Fig. 17.2. The masses are shown in atomic mass units and are all very close to 11 u, as one would expect. Nevertheless, it is clear that ^{11}B , with $Z = 5$ and $N = 6$, has the lowest mass and hence lowest rest energy. As one moves away from $Z \approx N$, the rest energies become progressively greater.

In heavier nuclei the Coulomb repulsion of the protons implies that the minimum occurs with Z appreciably less than N . This is illustrated in Fig. 17.3, which shows the masses of the $A = 231$ isobars, with the minimum at ^{231}Pa , for which $Z = 91$ and $N = 140$. In any case, the important point is that for any given set of isobars there is a definite value of Z , usually somewhat less than N , for which the rest energy is a minimum. Any nucleus with too many neutrons, such as ^{231}Th , or with too many protons, such as ^{231}U , has a rest energy larger than this minimum. (Sometimes there can be two minima; see Problem 17.25.)

Let us now consider a nucleus, such as ^{231}Th in Fig. 17.3, that has too many neutrons. That ^{231}Th has higher energy than ^{231}Pa would imply that ^{231}Th is unstable, provided that there were some mechanism to let ^{231}Th transform itself into the lower-energy ^{231}Pa ; and indeed, experiment shows that there is such a mechanism. The excess energy in ^{231}Th creates and ejects an electron, and one neutron is simultaneously converted into a proton. The resulting transformation,



is the process that Rutherford called β decay. Two such decays are indicated on Fig. 17.3, where they are labeled β^- to emphasize that they produce negative electrons.

A nucleus such as ^{231}Np , with too many protons, is also unstable. In fact, for most such nuclei there are two possible modes of decay. The excess energy in ^{231}Np can create and eject a positron (the positively charged version of the electron, mentioned briefly in Section 2.8) changing a proton into a neutron. The resulting transformation,

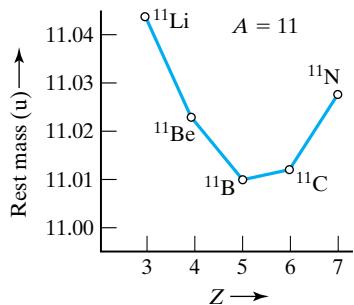
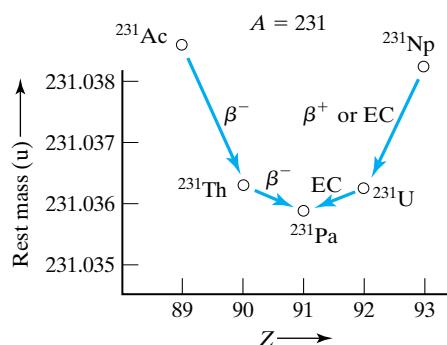


FIGURE 17.2

The rest masses, in atomic mass units, of the five most stable isobars with $A = 11$. Although all are close to 11 u, boron (with $Z = 5$ and $N = 6$) has the lowest mass and hence lowest rest energy.

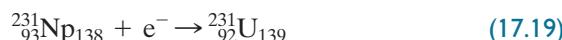
FIGURE 17.3

The rest masses of the isobars with $A = 231$ have their minimum at protactinium with $Z = 91$ and $N = 140$. The possible decays of the other species are indicated as β^- , β^+ , and EC (for electron capture), as described in the text.



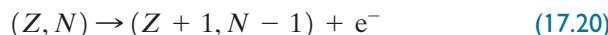
Section 17.4 • Beta Decay and the Neutrino 45

is called β^+ decay. Alternatively, the ^{231}Np nucleus can convert a proton into a neutron by capturing one of the atomic electrons:



a process called **electron capture**. Figure 17.3 indicates both of these processes for ^{231}Np . (Note in the same picture that ^{231}U decays by electron capture but not by β^+ emission. The impossibility of β^+ decay in ^{231}U is related to the slightly different energy requirements for these two processes. See Problem 17.23.)

The puzzling aspect of β decay, mentioned earlier, emerges when we consider the energy of the ejected electron or positron in β decay. Let us consider the case of β^- decay, for which the observed process can be written as



If the parent nucleus (Z, N) is at rest and if the process really involves only the two final bodies indicated in (17.20), then the energies of those two bodies are uniquely determined by conservation of energy and momentum. With the parent at rest, the total kinetic energy of the final particles is

$$K = \Delta Mc^2 = \{m_{\text{nuc}}(Z, N) - [m_{\text{nuc}}(Z + 1, N - 1) + m_e]\}c^2 \quad (17.21)$$

We can simplify this expression for ΔM by adding and subtracting Zm_e , to give

$$K = \Delta Mc^2 = \{m_{\text{atom}}(Z, N) - m_{\text{atom}}(Z + 1, N - 1)\}c^2; \quad (17.22)$$

that is, ΔM can be found by taking the difference of the masses of the two corresponding *atoms*.

Since the electron is so much lighter than the recoiling nucleus, the nucleus takes very little kinetic energy. Therefore, to an excellent approximation (about 1 part in 10^5 for a heavy nucleus — see Problem 17.34) all the kinetic energy (17.22) must go to the electron. We conclude that if the only particle ejected in β^- decay were the electron, as in (17.20), then all the electrons in a sample of β^- radioactive material (of one type) would emerge with exactly the same kinetic energy ΔMc^2 , as given by (17.22).

Example 17.3

Assuming that the electron's energy is given correctly by (17.22), find the kinetic energy of the electrons ejected in the β^- decay of bismuth 210 to polonium 210.

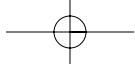
The atomic masses concerned are given in Appendix D as

$$m(^{210}\text{Bi}) = 209.984096 \text{ u}$$

$$m(^{210}\text{Po}) = 209.982848 \text{ u}$$

and the mass difference is therefore $\Delta M = 0.001248 \text{ u}$. Thus, according to (17.22), the electron's kinetic energy should be

$$K = \Delta Mc^2 = (0.001248 \text{ u}) \times \left(\frac{931.5 \text{ MeV}/c^2}{1 \text{ u}} \right) c^2 = 1.16 \text{ MeV}$$



46 Chapter 17 • Radioactivity and Nuclear Reactions

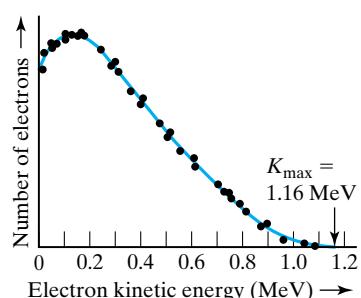


FIGURE 17.4

Distribution of kinetic energies of the electrons ejected in the β^- decay of ^{210}Bi . The observed values range from $K = 0$ up to a maximum denoted K_{\max} . It is found that K_{\max} is equal (or very nearly so) to the value 1.16 MeV predicted by (17.22) on the assumption that the electron is the only particle ejected.

Experiment shows that our prediction that all electrons should be ejected with the same energy is completely wrong. Figure 17.4 shows the distribution of energies of the electrons emitted by ^{210}Bi . Evidently, the energies range continuously from $K = 0$ up to the predicted value of 1.16 MeV (or very close to it). While a few electrons have nearly the predicted energy, the great majority have much less. This apparent failure of conservation of energy was beginning to be recognized in the late 1920s, and efforts were made to find the missing energy. For example, it was conceivable that some energy was carried away by undetected radiation. To check this possibility, β radioactive materials were enclosed in thick-walled calorimeters, but here, too, the apparent release of energy — that is, the heat produced in the calorimeter — was appreciably less than required by conservation of energy.

The suggestion was even made that conservation of energy might not apply to β decay. However, a less radical suggestion due to Pauli proved to be the correct explanation. In the early 1930s Wolfgang Pauli suggested that there might be another particle emitted in β decay, which interacted only very weakly with matter and hence could carry away the missing energy undetected. This suggestion would immediately explain the spread in the measured energies of the electrons: With two ejected particles (in addition to the residual nucleus) the released energy could be shared between the two particles in many different ways. For example, the electron could be left at rest while the hypothetical particle took all the available energy or vice versa. Thus the electron's kinetic energy could range anywhere from zero to some K_{\max} , as observed.

From the observed facts about β decay, we can immediately infer the main properties of Pauli's proposed particle — its charge, mass, and interaction with matter. First, its charge must be zero: The process (17.20) already conserves charge since the right-hand side has one extra charge of each sign (one proton and one electron). If there really is an additional particle on the right-hand side, conservation of charge requires that this additional particle be neutral.

Concerning the particle's mass, we saw in connection with Fig. 17.4 that the observed kinetic energies of the ejected electrons range close to the value ΔMc^2 predicted on the assumption that there was no additional particle. In other words, the total energy taken by the proposed additional particle varies *all the way down to zero* (or very nearly so). Since the minimum possible energy of a particle is its rest energy mc^2 , this requires that the new particle have mass very close to zero — probably less than an eV/ c^2 or so. Since the new particle has zero charge and very small mass, Fermi named it the **neutrino** (little neutral one), denoted by ν , the lowercase Greek letter nu.

Concerning its interactions, the great difficulty of detecting the neutrino implies that it interacts very weakly with all matter. In fact, we now know that, like electrons, neutrinos are completely immune to the strong nuclear force; in addition, they are immune to all electromagnetic forces. The one force to which neutrinos are subject is the so-called weak force, which we discuss in Chapter 18 (and also presumably gravity, but this is usually negligible in nuclear processes). For this reason, the neutrino was not directly detected until 1956. Nevertheless, the faith in conservation of energy was so strong that well before then, most physicists confidently believed that the neutrino would eventually be found. This belief was reinforced by the realization that if there were no neutrino, β decay would also violate the conservation laws of momentum (Problem 17.24) and angular momentum.

Section 17.4 • Beta Decay and the Neutrino 47

It eventually became clear that there are two closely related particles, the neutrino ν and the antineutrino $\bar{\nu}$, and it is actually the antineutrino that is emitted in β^- decay:

$$(Z, N) \rightarrow (Z + 1, N - 1) + e^- + \bar{\nu}. \quad (17.23)$$

On the other hand, it is the neutrino that emerges in β^+ decay,

$$(Z, N) \rightarrow (Z - 1, N + 1) + e^+ + \nu \quad (17.24)$$

and also in electron capture (see Problem 17.23). We will discuss the difference between the ν and $\bar{\nu}$ in Chapter 18. Even though the two particles are different, it is common practice (which we follow) to use the one name “neutrino” for both ν and $\bar{\nu}$ in any context where the distinction is unimportant.*

Instability of the Neutron

Before we describe the discovery of the neutrino, we must mention briefly one particularly striking example of β decay: the decay of the neutron. A glance at the masses inside the front cover shows that the rest mass of the neutron is greater than that of the proton by enough to create the rest energy of an electron. This means that the neutron is unstable and can decay via the β^- decay

$$n \rightarrow p + e^- + \bar{\nu} \quad (17.25)$$

This rather surprising prediction is borne out by experiment: An isolated neutron is indeed unstable and decays as in (17.25) with a half-life of about 10 minutes.

This immediately raises the question: If the neutron is unstable, how is it that normal stable matter contains large numbers of neutrons? To answer this, consider again the $A = 11$ isobars shown in Fig. 17.5. A nucleus like ^{11}Be , with too many neutrons, is indeed unstable, because the total energy of ^{11}Be is greater than that of ^{11}B , and it is energetically favorable for one of the neutrons in ^{11}Be to undergo β^- decay, changing ^{11}Be to ^{11}B . If we look instead at ^{11}B , we find a different story: Because of the properties of the nuclear force and the Pauli principle, the next nucleus over, ^{11}C , has *higher* total energy. Therefore, it is not energetically favorable for a neutron in ^{11}B to convert to a proton, and ^{11}B is stable. The situation is clearly this: Even though the isolated neutron is β^- unstable, a nucleus (Z, N) may or may not be β^- unstable, depending on whether its neighboring isobar $(Z + 1, N - 1)$ has less or more rest energy.

If we go farther to the right in Fig. 17.5 and look at ^{11}C , we find that here the balance has tipped against the protons: In this case it is energetically favorable for a *proton* to change into a neutron. More generally, in any atom (Z, N) with more rest energy than its neighbor $(Z - 1, N + 1)$ it is energetically favorable for a proton to convert to a neutron, and such atoms are unstable against β^+ decay or electron capture.

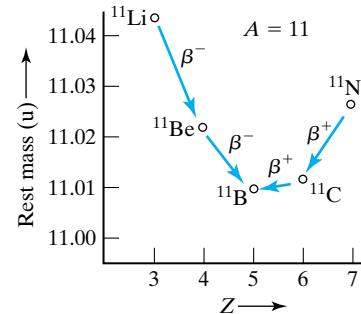
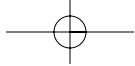


FIGURE 17.5

The neutrons in ^{11}Be and ^{11}Li are unstable; that is, it is energetically favorable for a neutron to change into a proton via β^- decay. In ^{11}C and ^{11}N it is the protons that are unstable. In ^{11}B both neutrons and protons are stable. (Both ^{11}C and ^{11}N can also decay by electron capture.)

*As we describe in Chapter 18, there are now known to be three kinds of neutrinos, each with its own antiparticle. Only one of these, the so-called electron neutrino, will concern us here.



48 Chapter 17 • Radioactivity and Nuclear Reactions

Observation of the Neutrino

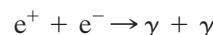
The difficulty of observing the neutrino is that it interacts so weakly with matter. Theoretical arguments indicate that if the β^- decay (17.25) is possible, so also is the so-called *inverse β* decay,



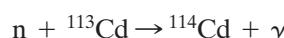
in which a $\bar{\nu}$ collides with a proton and converts it to a neutron plus positron. If an antineutrino from a β^- decay were to pass through a material containing protons (such as water), it could, in principle, cause the reaction (17.26) and hence be detected. Unfortunately, this reaction, although possible, is exceedingly improbable. It has been calculated, for example, that on average an antineutrino would travel several hundred light years through water before producing a single inverse β decay.

Nevertheless, the reaction (17.26) was observed by the American physicists Clyde Cowan and Frederick Reines in 1956. Their first requirement was obviously to have a source of an enormous number of antineutrinos. For this, they chose a nuclear fission reactor. As we discuss in Section 17.7, reactors produce large quantities of material that is β^- radioactive and hence generate many antineutrinos. Their detector, a large tank of water, was placed close to a reactor where, they calculated, some 10^{20} antineutrinos would pass through it each hour. The product of this enormous number of antineutrinos and the tiny probability for each $\bar{\nu}$ to induce the reaction (17.26) gave their expected number of events — about 3 per hour.

To register the occurrence of the reaction (17.26), they needed some way to detect the newly created positron and neutron. The positron rapidly slows down in the water and, within about a nanosecond, annihilates with an atomic electron to give two photons, each of about 0.5 MeV:



(see Section 2.8). The neutron moves off more slowly. To detect its presence, Cowan and Reines added cadmium chloride to the water, since the cadmium nucleus has a high probability of capturing neutrons. Within about a microsecond, the neutron is captured and the cadmium gives off a photon (or several photons) of total energy about 9 MeV:



Thus, both the positron and neutron eventually signal their presence by producing γ -ray photons of known characteristic energies, as shown in Fig. 17.6(a).

The γ -rays were detected by sandwiching the water tank between large tanks of scintillator, a fluid that gives off visible light when traversed by γ -rays, as shown in Fig. 17.6(b). The visible light was detected by photocells surrounding the scintillator tanks, and in this way both the timing and energy of the γ -rays could be recorded. Finally, the entire apparatus was surrounded by thick shielding (mostly lead and paraffin) to keep out any particles other than neutrinos.

After more than 1000 hours of observation and numerous consistency checks, Cowan and Reines were able to announce the positive identification of the antineutrino by verifying the inverse β decay (17.26) with the characteristic time sequence of photons that it produced.

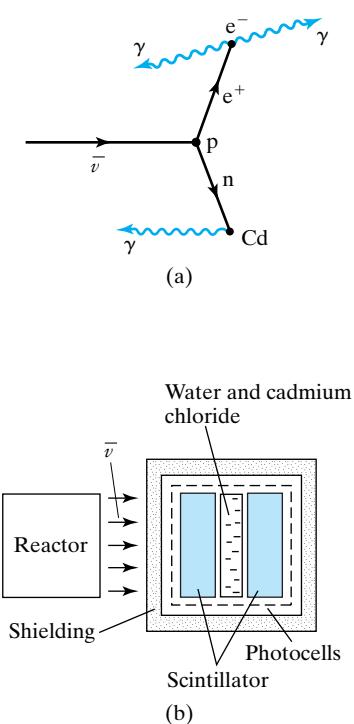


FIGURE 17.6

(a) Sequence of events for detecting the inverse β decay, $\bar{\nu} + p \rightarrow e^+ + n$. The positron annihilates with an electron to give two photons; the neutron is captured by a cadmium nucleus, which emits one or more photons. All of these events take place in the water tank; when the photons pass through the neighboring tanks of scintillator, they produce visible light, which is detected by photocells. **(b)** Arrangement of the reactor (which produces the neutrinos) and the tanks of water and scintillator. The shielding excludes all particles except the neutrinos, but is completely transparent to neutrinos.

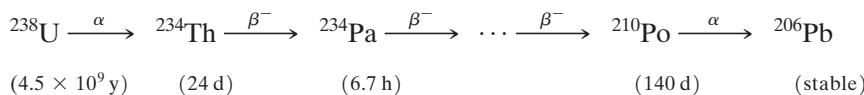
17.5 The Natural Radioactive Series

Most of the naturally occurring radioactive nuclei are heavy nuclei, with $A \geq 210$, that can be grouped into three radioactive series. These three series owe their existence to the occurrence of three parent nuclei with exceptionally long half-lives:

Parent	Half-Life	
^{232}Th	14 billion years	
^{235}U	0.70 billion years	
^{238}U	4.5 billion years	(17.27)

(where 1 billion = 10^9). These nuclei are believed to have been formed, along with many other radioactive nuclei, in the nuclear processes of nearby supernova explosions shortly before the birth of the solar system. When the solar system condensed, it contained some of this radioactive material, most of which has decayed in the 5 billion or so years since that time. However, because the three half-lives (17.27) are of order 1 billion years, the original supply of these three nuclei is still not exhausted.*

When nuclei of the three species (17.27) decay, they give birth to radioactive “children,” which in turn produce radioactive grandchildren, and so on. This process continues, producing three radioactive series, or chains, which stop only when a stable descendant is reached. For example, the series that begins with ^{238}U ends at ^{206}Pb . Some of the stages in between (with half-lives shown in years, days, or hours) are as follows:

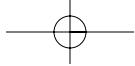


Notice that the half-lives of the descendants can be, and are, much shorter than that of the original parent. The natural occurrence of ^{234}Th , for example, does not depend on its having survived for 5 billion years; rather, it is being produced continually by its long-lived parent ^{238}U .

Notice also that both α and β decays occur in the same radioactive series. In addition, the offspring nuclei are often produced in excited states that de-excite by γ decay. Thus, all three types of decay will be observed in any natural sample. Further, a natural sample of any of the three original parents is bound to contain a mixture of all the descendants, each decaying at its own characteristic rate. Under the circumstances it is remarkable that the early workers — the Curies, Rutherford, and others — were able to unravel what was happening as quickly as they did.

To better understand the sequence of decays in the three radioactive series, let us consider a plot of Z and N for the stable nuclei. This *curve of stability* was shown in Fig. 16.1 and is reproduced more schematically in Fig. 17.7(a). If we imagine rest masses plotted upward, out of the paper, stable nuclei would be lower than neighboring unstable nuclei, and the curve in Fig. 17.7(a) would become the floor of a **valley of stability**, as sketched in Fig. 17.7(b). Nuclei on the sloping sides of this valley, to the right or left of the

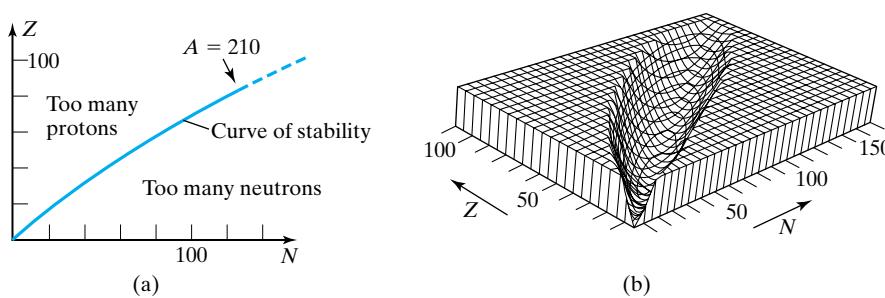
*Notice, however, that the half life of ^{235}U is much shorter than that of ^{238}U . This is why the isotope ^{235}U is much less abundant today. See Problem 17.28.



50 Chapter 17 • Radioactivity and Nuclear Reactions

FIGURE 17.7

(a) The curve of stability is the locus of the most stable nuclei in a plot of Z against N . For $A \geq 210$, even those nuclei on the curve are unstable. (b) The valley of stability. The variables plotted horizontally are Z and N , exactly as in (a). The variable plotted upward was chosen to be the mass excess, $m - A$, since this excludes the automatic, and hence uninteresting, increase of m as A increases. (The plot was truncated at a mass excess of $100 \text{ MeV}/c^2$ to show the valley more clearly.) The curve in (a) can be seen as an aerial view of the valley floor in (b).



curve in Fig. 17.7(a), have too many neutrons or too many protons and they decay by β^- or β^+ emission.

For heavy nuclei, the increasing Coulomb repulsion of the protons raises the valley floor as A increases, and even those nuclei on the valley floor are unstable for $A \geq 210$. It turns out that the most favorable decay for such nuclei is α decay, as the following example illustrates.

Example 17.4

Use the measured masses in Appendix D to find the kinetic energy released in the α decay of ^{232}Th at rest



Compare this with the kinetic energy released in the hypothetical neutron decay



The atomic masses involved in the α decay are (Appendix D)

Initial	Final
$m(^{232}\text{Th}) = 232.0381 \text{ u}$	$m(^{228}\text{Ra}) = 228.0311 \text{ u}$
	$m(^4\text{He}) = 4.0026 \text{ u}$
Total $M_i = 232.0381 \text{ u}$	Total $M_f = 232.0337 \text{ u}$

Therefore, the kinetic energy released is

$$K = (M_i - M_f)c^2 = (0.0044 \text{ u}) \times \frac{931.5 \text{ MeV}}{1 \text{ u}} = 4.1 \text{ MeV}$$

Strictly speaking, the decay (17.28) is a nuclear process, and we should have used the masses of the three *nuclei*, rather than the three atoms. However, we get the same answer using the atomic masses because there are equal total numbers of atomic electrons on both sides of (17.28), and their masses cancel when we take the difference $M_i - M_f$.

The corresponding masses in the hypothetical neutron decay (17.29) are

Initial	Final
$m(^{232}\text{Th}) = 232.0381 \text{ u}$	$m(^{231}\text{Th}) = 231.0363 \text{ u}$
	$m_n = 1.0087 \text{ u}$
Total $M_i = 232.0381 \text{ u}$	Total $M_f = 232.0450 \text{ u}$

Section 17.5 • The Natural Radioactive Series 51

We see that the combined rest mass of the final bodies is greater than the initial mass, and the proposed neutron decay (17.29) cannot occur.

The reason that α decay is so highly favored is the exceptionally large binding energy of the alpha particle. In each of the decays (17.28) and (17.29) the binding energy per nucleon (B/A) of the offspring nucleus is greater than that of the parent. However, in (17.29) the ejected neutron has no binding energy at all, and the net effect of the decay proves to be a loss of binding energy — or a gain of rest energy — and the process is forbidden. In (17.28), on the other hand, the large binding energy of the alpha particle results in a net gain of binding energy, and the process is energetically favored. The same considerations apply to all unstable nuclei near the floor of the valley of stability: Alpha decay is the favored, and often the only, mode of decay.

When a nucleus emits an alpha particle, it loses two protons and two neutrons. This moves it in a 45° , “southwest” direction (that is, down and to the left) on the plot of Z against N , as shown in Fig. 17.8(a). If the parent is on the valley floor, this means that the offspring will be on the neutron-rich side of the valley. If the offspring is far enough up the valley side, it will undergo β^- decay to move it northwest back to the valley floor, as in Fig. 17.8(b). Otherwise, it will undergo further α decays, each of which carries it farther up the neutron-rich side of the valley until, finally, one or more β^- decays take it back to the valley floor, as in Fig. 17.8(c).

We see that a radioactive series will consist of a succession of α and β^- decays (plus γ decays when an offspring is formed in an excited state), which zigzag down the valley of stability. Because of the orientation of the valley floor, the path never strays appreciably onto the proton-rich side of the valley. Therefore, β^+ decays (and electron capture) have no part in the natural radioactive series. The whole of the series that begins with ^{232}Th is shown in Fig. 17.9.

At the start of the series shown in Fig. 17.9 an α decay is followed immediately by two β^- decays. The effect of these three decays is that N decreases by 4 while Z returns to its original value; that is, the fourth nucleus (^{228}Th) is an isotope of, and hence chemically indistinguishable from, the first (^{232}Th). The presence of different isotopes of several elements in the radioactive series was a source of confusion to the early nuclear physicists, but eventually led to the

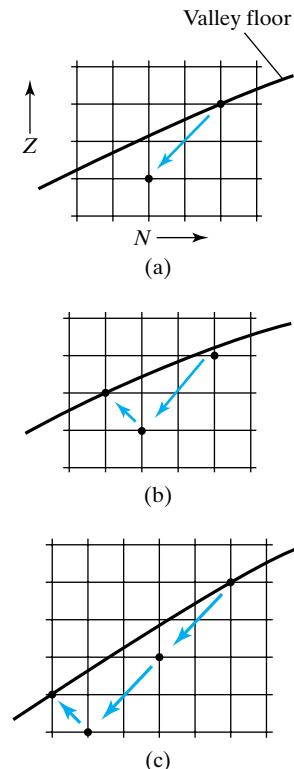


FIGURE 17.8

- (a) In α decay both Z and N decrease by 2, and the offspring nucleus lies exactly southwest of the parent. If the parent is on the valley floor, this puts the offspring on the neutron-rich side of the valley.
- (b) If the offspring is high enough on the valley side, it will undergo β^- decay, producing a grandchild nearer the valley floor.
- (c) Depending on the position and orientation of the valley floor, a series may make two or more α decays before returning to the valley floor by β^- decay.

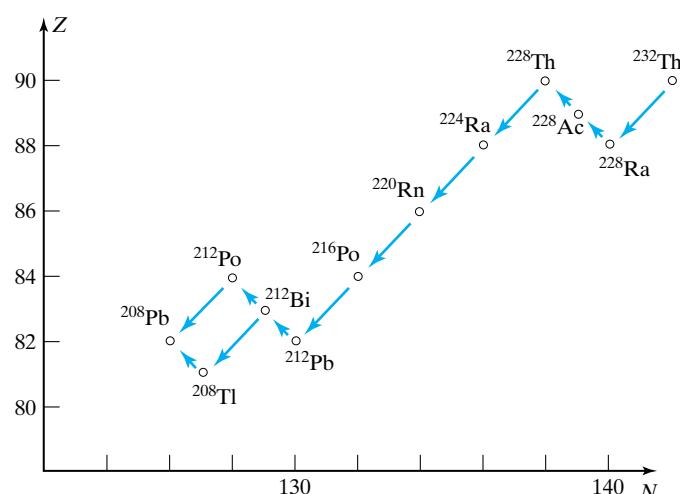
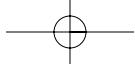


FIGURE 17.9

The radioactive series that originates in ^{232}Th follows a zigzag path, staying as close as possible to the floor of the valley of stability. Notice the branch at ^{212}Bi .



52 Chapter 17 • Radioactivity and Nuclear Reactions

first clear understanding that there can be several different nuclei, all belonging to the same chemical element.

Notice also in Fig. 17.9 the branch at ^{212}Bi , which can undergo both α and β^- decay. When this happens, we cannot predict which decay will actually occur; rather, we can state only the probabilities of the two modes: 64% for β^- and 36% for α , in the case of ^{212}Bi .

To conclude this section, it is perhaps worth emphasizing the great importance of the natural radioactive series, both in the history of nuclear physics and in many other fields today. The discovery of natural radioactivity in 1896 was the beginning of nuclear physics. Rutherford's identification of the atomic nucleus in 1911 used alpha particles from the radioactive series, and the predicted energy dependence of Rutherford scattering was verified by using alpha particles of different energies from different sources. As we have just seen, the important concept of isotopes emerged from the study of radioactive series. As we describe in the next section, alpha particles from the radioactive series produced the nuclear reactions that established the presence of protons in nuclei and the existence of the neutron. The invention in the early 1930s of machines that could accelerate charged particles sharply reduced the importance of natural alpha particles as nuclear probes. Nevertheless, alpha particles from radioactive materials are still used to make convenient sources of neutrons, and such sources were used to produce artificial elements and to induce nuclear fissions.

Today natural radioactivity is used in several ways to study the earth's crust. The measured ratios of parent to offspring atoms in a rock can date its solidification. An important member of the radioactive series is radon, a noble gas that moves easily through porous rock because it is chemically inactive and not easily absorbed.* Because radon is radioactive, even minute quantities seeping out of the earth's surface can be detected quite easily. Sudden increases in radon levels have been associated with the rock fractures preceding earthquakes. Also, the principal cause of the slow movement of the earth's crust appears to be convection currents that are driven by heat from the natural radioactivity distributed throughout the earth's volume.

Perhaps most important in our everyday lives, radioactivity plays an enormous role in modern medicine. Radiation from radioactive nuclei (and other sources), when carefully directed at cancerous tissue, can kill the diseased cells and is one of our best weapons against cancer. (The unavoidable damage done to neighboring healthy cells causes the notorious side effects of this therapy.) It is sometimes possible to place radioactive isotopes directly in a tumor, where it will slowly kill the cancer; for example, radioactive iodine, ^{131}I , when injected into the bloodstream tends to concentrate in the thyroid and can be used to treat cancer of the thyroid. Radioactive tracers are widely used in medical research. A compound synthesized with a radioactive nucleus can be traced as it moves around the body in the blood or in food, for example. Compounds made with β^+ emitters, chosen so that they concentrate in a region of interest, will emit positrons whose prompt annihilation with nearby electrons emits γ -rays that can be detected and used to create an image in the process called *positron emission tomography* or PET.

*For this reason, radon can seep into buildings, where it is a potential health hazard since it can be inhaled into the lungs. Those radons that decay while in the lungs produce the chemically active polonium, which sticks to the lung walls, so that all its subsequent decays irradiate the nearby tissue.

17.6 Nuclear Reactions

We have so far discussed nuclear transformations that result from the spontaneous decay of a single nucleus. While there is much to learn from such processes, there is even more to learn from **nuclear reactions**, the transformations that occur when two nuclei collide. The first such reaction was observed by Rutherford in 1919 when he bombarded nitrogen nuclei with alpha particles (from a natural radioactive source) and produced the reaction



There is an obvious parallel between a nuclear reaction such as (17.30) and a chemical reaction, for example

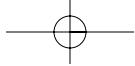


in which molecules (or atoms and molecules) collide and rearrange their constituent atoms. However, there is an important practical difference: Under normal conditions on earth, chemical reactions occur naturally and abundantly, whereas nuclear reactions occur hardly at all. The reason for this difference is easy to see. For two atoms or molecules to react chemically, they only need to approach one another closely enough for their outer electrons to overlap, and this condition is easily met at normal temperatures and densities. But even in the most violent chemical reactions the nuclei, buried deep inside their protective shells of electrons, remain very far apart compared to the nuclear force range. Even if two nuclei were stripped of their electrons, a large energy would still be needed to bring them close together, because of the Coulomb repulsion between their positive charges. For example, the energy needed to bring a proton to the surface of an alpha particle is about 1 MeV (Problem 17.38). For this reason, nuclear reactions occur naturally only very rarely on earth.* On the other hand, they do occur — and are the main source of energy — in stars, where the temperature is so high that the kinetic energy of thermal motion is sufficient to overcome the Coulomb repulsion.

To produce a nuclear reaction in the laboratory, one must first produce nuclei with energies of order 1 MeV or more and then direct them at other nuclei (contained in a suitable solid, liquid, or gas). The energy of the projectiles lets them sweep through the electrons of a target atom, with some of them approaching the nucleus close enough to induce a nuclear reaction. Until the early 1930s the only available projectiles were the alpha particles from natural radioactive sources. Then, two developments resulted in several new possibilities. First, the development of particle accelerators made it possible to accelerate many different nuclei to an energy sufficient to induce nuclear reactions. Second, soon after the discovery of the neutron, it was realized that because they are electrically neutral, even low-energy neutrons can approach and penetrate a nucleus. The study of nuclear reactions induced by low-energy neutrons quickly became a popular and fruitful area of research.

In this section we describe just a few nuclear reactions of historical importance, starting with the reaction (17.30). In Sections 17.7 and 17.8 we discuss the special cases of fission and fusion.

*There are a few examples, mostly induced by high-energy cosmic-ray particles, for instance the reaction $\text{n} + {}^{14}\text{N} \rightarrow \text{p} + {}^{14}\text{C}$ that produces carbon 14 in the atmosphere.



54 Chapter 17 • Radioactivity and Nuclear Reactions

Identification of the Proton as a Nuclear Constituent

The first nuclear reaction to be observed was the reaction (17.30), which we can rewrite in more detail as



This reaction was induced by the alpha particles from ${}^{212}\text{Bi}$ as they passed through air. Rutherford noticed that particles were being produced with a range in air much longer than that of the alpha particles, but similar to that of hydrogen nuclei of comparable energy. He immediately suspected (and eventually proved) that these particles were hydrogen nuclei ejected from nitrogen nuclei in the air. He then argued that hydrogen nuclei are presumably contained in all nuclei and provide the positive charge and much of the nuclear mass. It was for this reason that he proposed a special name for the hydrogen nucleus: namely, proton (Greek for “first one”).

At about this time the British physicist C. T. R. Wilson developed the cloud chamber, in which moving charged particles leave tracks of condensed water droplets. In this way it became possible to photograph the tracks of particles involved in nuclear reactions, as in Fig. 17.10, which shows a cloud chamber photograph of reaction (17.32).

We can use the reaction (17.32) to illustrate two conservation laws that apply to all nuclear reactions. Consider first the mass number, or nucleon number, A , which appears as a superscript on each symbol. If you add up the superscripts on either side of (17.32), you will see that the total number of nucleons is the same before and after the reaction. Thus,

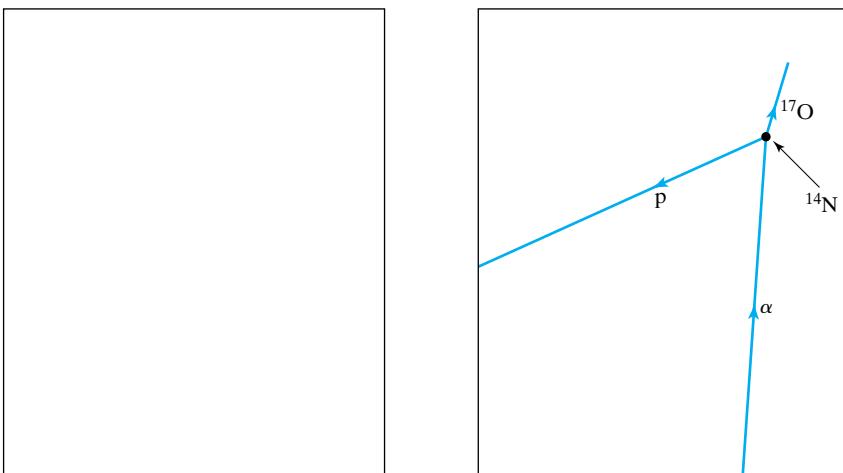
$$\sum A_{\text{initial}} = 4 + 14 = 18$$

and

$$\sum A_{\text{final}} = 1 + 17 = 18$$

FIGURE 17.10

An early cloud chamber photograph of the reaction $\alpha + {}^{14}\text{N} \rightarrow p + {}^{17}\text{O}$. Numerous alpha particles, from ${}^{212}\text{Bi}$, enter the picture from below. One of them strikes a nitrogen nucleus (stationary and hence invisible) and produces a long-range proton and an oxygen nucleus. Note that the source produced alpha particles of two different energies; those with low energy all stop less than halfway up the picture; those with high energy almost all reach the top.



and hence

$$\sum A_{\text{initial}} = \sum A_{\text{final}} \quad (17.33)$$

It is found that this equality holds in all nuclear reactions; the total number of nucleons does not change. We refer to this observation as the **law of conservation of nucleon number.*** It is a useful aid in checking whether a proposed reaction is possible: Any nuclear reaction that does not satisfy (17.33) is impossible.

A second conservation law that applies to all reactions is the familiar law of conservation of charge; the total charge cannot change.

$$\sum q_{\text{initial}} = \sum q_{\text{final}} \quad (17.34)$$

In (17.32) the only charged particles are protons, and conservation of charge implies simply that the total number of protons (shown as a subscript for each nucleus) cannot change.

$$\sum Z_{\text{initial}} = 9 = \sum Z_{\text{final}}$$

In general, however, the number of protons can change in processes that involve other charged particles as well. For example, in the β decay



the number of protons increases by one, but this increase of charge is balanced by the production of the negative electron. The law of conservation of charge is another useful aid in checking whether a given process can occur.

Discovery of the Neutron

The neutron was discovered by Rutherford's colleague James Chadwick in the reaction



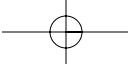
(You should check quickly that this satisfies the two conservation laws discussed in the last two paragraphs.) This reaction had actually been studied by the French physicists Irène and Frédéric Joliot-Curie, who had found that when alpha particles strike beryllium, a neutral “radiation” is produced, which can eject protons from any material containing hydrogen, such as paraffin wax (Fig. 17.11). Since the neutron was not yet known, the Joliot-Curies made the natural suggestion that the neutral radiation consisted of photons. However, as Chadwick emphasized, protons were ejected from the wax with momenta up

Irene and Frédéric Joliot-Curie
(1897–1956, French)



Irène Joliot-Curie was the daughter of Pierre and Marie Curie. She met her husband, Frédéric Joliot, when they were both working as assistants to her mother. After they were married, they took the name Joliot-Curie and continued to work together. They narrowly missed discovering the neutron (losing out to Chadwick) and the positron (losing out to Anderson). In 1934 they identified the first man-made radioactive nucleus, an isotope of phosphorus, which they created by bombarding aluminum with alpha particles; and for this discovery, they won the Nobel Prize for chemistry in 1935. Irène was refused admission to the American Chemical Society in 1954 because of her left-leaning politics.

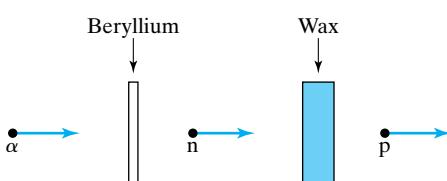
*As we will see in Chapter 18, this law has to be generalized. Nucleons are just two members of a whole class of particles called *baryons*, and it is actually the number of baryons that cannot change. However, in most nuclear reactions the only baryons involved are neutrons and protons, and (17.33) therefore holds.



56 Chapter 17 • Radioactivity and Nuclear Reactions

FIGURE 17.11

Discovery of the neutron. An unknown neutral particle, now known to be the neutron, was produced when alpha particles collided with beryllium nuclei. The particle's presence was evident since it could knock protons out of a block of wax; that it was neutral was clear because it produced so little ionization.



James Chadwick (1891–1974, English)



James Chadwick had worked with Rutherford in Manchester, but was unlucky enough to be studying on a scholarship in Germany when World War I broke out, and he spent the war in an internment camp. After the war he moved to Cambridge and worked again with Rutherford. In 1932 he picked up on work of the Joliot-Curies and showed that bombardment of beryllium by alpha particles produced neutrons, a discovery which won him the Nobel Prize for physics in 1935.

to about $100 \text{ MeV}/c$. It is easy to show (Problem 17.45) that if a photon strikes a proton and gives it $100 \text{ MeV}/c$ of recoil momentum, the photon itself must have had a momentum of at least $50 \text{ MeV}/c$, and hence an energy of 50 MeV . (Recall that $E = pc$ for a photon.) But it is hard to see how photons with this much energy could possibly have been produced when the beryllium was struck by the original alpha particles, whose energy was only about 5 MeV .

Chadwick proposed instead that the “radiation” consisted of neutral particles, which he called neutrons,* whose mass was close to that of the proton. This suggestion immediately solved the problem of the photon energy: In the collision of two *equal* masses, the target particle (the proton) can take *all* the kinetic energy of the incident projectile (the neutron). Thus, the neutrons needed to have only the kinetic energy of the recoil protons, which, for $p = 100 \text{ MeV}/c$, was only about 5 MeV ($K = p^2/2m \approx 5 \text{ MeV}$), and this was entirely reasonable.

In a series of experiments Chadwick let neutrons from the same source strike various different nuclei. By measuring the recoil energy of the different targets, he was able to estimate the neutron’s mass as about 1.15 u . For a more accurate measurement, he used the mass-energy relation and the known masses of the nuclei involved in the neutron-producing reaction (Problem 17.39). He concluded that the neutron mass “probably lies between 1.005 and 1.008,” in remarkable agreement with the current value of 1.009 u .

Reactions Produced by Man-Made Accelerators

As we describe in Section 17.11, the early 1930s saw the development of several *accelerators* — machines that accelerate charged particles, at first to energies of order 10^5 eV and, today, up to 10^{12} eV and more. The first nuclear reaction produced by an accelerator was the reaction



which was observed in 1932 by John Cockcroft and Ernest Walton (both associates of Rutherford) using 600-keV protons from an accelerator of their own design.

We can use the reaction (17.36) to illustrate another of the conservation laws that apply to all reactions — the conservation of relativistic energy. Since the energy of each body in (17.36) has the form $K + mc^2$, conservation of energy implies that

$$K_i + M_i c^2 = K_f + M_f c^2$$

*The idea of the neutron, and the name, had been suggested several years earlier by Rutherford.

where K_i and K_f denote the initial and final total kinetic energies in (17.36) and M_i and M_f are the corresponding total rest masses. Therefore, the total gain in kinetic energy is just the total loss of rest energy.

$$\Delta K = K_f - K_i = (M_i - M_f)c^2 \quad (17.37)$$

(The gain ΔK of kinetic energy in a reaction is often called the **Q value** and denoted by Q .) Cockcroft and Walton were able to measure ΔK and got the value 17.2 MeV, in satisfactory agreement with the modern value 17.35 MeV for $(M_i - M_f)c^2$. In this way they gave the first experimental verification of Einstein's mass-energy relation.

Neutron-Induced Reactions

Because the neutron is electrically neutral, it can easily approach and penetrate nuclei, even at very low energies. One of the first to exploit the neutron as a probe of nuclei was the Italian physicist Enrico Fermi. As a source of neutrons, he used a mixture of a radioactive α emitter and powdered beryllium, in which the alpha particles produced neutrons by Chadwick's reaction (17.35). (This arrangement is still used today as a convenient portable source of neutrons.) Fermi systematically fired the neutrons at about 70 of the known elements and produced a variety of nuclear reactions and radioactive products. Most of the reactions that he observed were among the following three types:



In all three cases the neutron-to-proton ratio N/Z increased, and most of the resulting nuclei were β^- radioactive. As we describe in the next section, another reaction that Fermi produced was the neutron-induced fission of uranium, but this reaction went unrecognized for a few more years.

Enrico Fermi
(1901–1954, Italian-American)



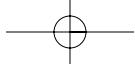
Fermi was a great theorist, an innovative experimentalist, and an inspiring teacher. While a professor at Rome he made important contributions to the theory of beta decay and carried out an experimental study of neutron-induced reactions, which led ultimately to the discovery of fission. After receiving the Nobel Prize in 1938 (for the neutron work), he fled from Fascist Italy to the United States, where he soon began work on neutron-induced fission. This resulted in the first controlled nuclear chain reaction, which was a key step in the development of nuclear weapons.

Cross Sections

It is important to be able to characterize the probabilities (either measured or calculated) of different nuclear reactions. This is traditionally done with a parameter called the *reaction cross section*, which is defined as follows: We imagine first a single target particle [for example, the nucleus (Z, N) in the reaction (17.38)]. We now send in a beam of projectiles [such as the neutron in (17.38)] such that a number n_{inc} of projectiles are incident per unit area perpendicular to the incident direction:

$$n_{\text{inc}} = \text{flux of projectiles} \quad (\text{particles/area})$$

Now, let N denote the number of reactions that occur (of the particular kind in question). As one might expect, N is proportional to the flux, n_{inc} , of



58 Chapter 17 • Radioactivity and Nuclear Reactions

projectiles, and we define the **reaction cross section** (or just **cross section**) σ as the ratio of these two quantities:

$$\sigma = \frac{N}{n_{\text{inc}}} = \frac{\text{number of reactions}}{\text{flux of projectiles}} \quad (17.41)$$

The cross section σ is large for reactions that are very probable and small for those that are improbable. Since σ has the dimensions of an area, we can think of it as the size of the target as “seen” by the projectile. In fact, if we imagine a classical experiment in which the incident particles are small pellets and the target is just a flat disc, and if we count each hit as a “reaction,” it is clear that

$$(\text{number of reactions}) = (\text{flux of pellets}) \times (\text{area of target})$$

Therefore,

$$(\text{area of target}) = \frac{\text{number of hits}}{\text{flux of pellets}}$$

Comparing this with (17.41), we see that in this classical situation, *the cross section σ is precisely the cross-sectional area of the target*. In the nuclear case, we can think of the cross section as the *effective area* of a target particle for reacting with a particular projectile. For example, the cross section for the inverse β decay $\bar{\nu} + p \rightarrow e^+ + n$ (a very improbable reaction) is found to be about $6 \times 10^{-48} \text{ m}^2$. We can therefore say that, as seen by a neutrino, the proton is exceedingly small, in the sense that almost all neutrinos will pass it unaffected.

For most nuclear reactions, the cross section is found to be of the order of the geometrical size of nuclei. (This is what you would expect in any situation where our classical analogy makes sense.) Since typical nuclear dimensions are of order 10^{-14} m , a natural unit of area for cross sections is 10^{-28} m^2 , and this unit has come to be called the **barn**.*

$$1 \text{ barn} = 10^{-28} \text{ m}^2 \quad (17.42)$$

For example, the cross section for 10-keV neutrons to be captured by ^{238}U is about 0.7 barn. On the other hand, the cross section for 0.1-eV neutrons to be captured by ^{113}Cd is unusually large, nearly 10^5 barns. As “seen” by low-energy neutrons, a cadmium nucleus is extremely big, in the sense that the probability of their capture is very large. This is why cadmium was used by Cowan and Reines to catch neutrons, as described in Section 17.4.

Once a reaction cross section σ has been measured or calculated, one can use it to predict the number of reactions that will occur in a given situation. We can solve (17.41) for the number of reactions produced by a single target nucleus.

$$N(\text{for one target nucleus}) = \sigma n_{\text{inc}} \quad (17.43)$$

In practice, all targets contain many nuclei, and we must multiply (17.43) by the total number, N_{tar} , of target nuclei to give the total number of reactions.

$$N = \sigma n_{\text{inc}} N_{\text{tar}} \quad (17.44)$$

*As in “it’s as big as a barn.”

Example 17.5

In the Cowan and Reines experiment, their water tank contained about 3×10^{28} protons ($N_{\text{tar}} = 3 \times 10^{28}$). In each hour the flux of neutrinos from the reactor was about $n_{\text{inc}} = 1.7 \times 10^{19} \text{ m}^{-2}$. Given that the cross section for inverse β decay is 6×10^{-20} barn, how many events could they expect in each hour?

According to (17.44), the number of events should have been

$$\begin{aligned} N &= \sigma n_{\text{inc}} N_{\text{tar}} \\ &= (6 \times 10^{-20} \times 10^{-28} \text{ m}^2) \times (1.7 \times 10^{19} \text{ m}^{-2}) \times (3 \times 10^{28}) = 3 \end{aligned}$$

which is what they observed.

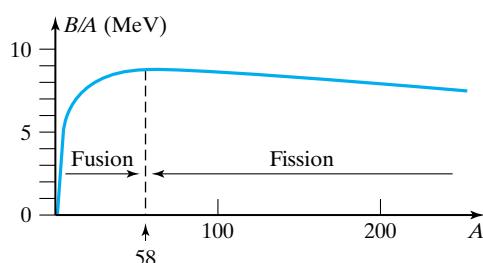
17.7 Nuclear Fission

In many nuclear processes there is a large release of kinetic energy; for example, in typical α decays, 4 to 9 MeV is released. Nuclear fission and fusion are two processes in which the large energy produced can be put to practical use on a macroscopic scale. If the energy is released rapidly, it can cause a violent explosion, and if it is released more slowly, it can be used to boil water and drive a steam turbine to generate electricity.

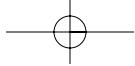
To understand what fission and fusion are and why they produce so much energy, we consider again the graph of the nuclear binding energy per nucleon, B/A , as a function of the mass number A . This graph was shown in Fig. 16.14, which we reproduce here as Fig. 17.12. It shows that B/A increases with A , up to a maximum around $A = 58$, and then decreases steadily all the way to the heaviest nuclei. This means that if we combine two light nuclei, the resulting single nucleus will have higher total binding energy. Therefore, the fusion of two light nuclei, for example,



would release energy, as we discuss in Section 17.8. On the other hand, when a heavy nucleus splits into two pieces, for example,

**FIGURE 17.12**

The binding energy per nucleon, B/A , as a function of A . Both the fusion of light nuclei and the fission of heavy nuclei lead to a higher binding energy and hence to a release of kinetic energy.



60 Chapter 17 • Radioactivity and Nuclear Reactions

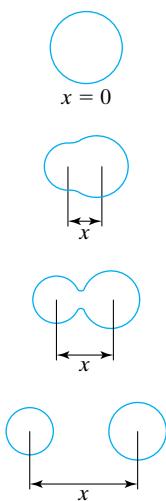


FIGURE 17.13

Excitation of a nucleus can cause it to change shape, develop a neck, and fission into two separate pieces. The separation can be characterized by the center-to-center distance x .

the total binding energy of the two fragments will be greater than that of the single parent, and again there will be a release of kinetic energy. It is this process of **nuclear fission*** that we discuss first.

We can easily estimate the energy released in a nuclear fission such as (17.46). From Fig. 17.12 we see that B/A for the two fragments with ($A \approx 100$) is nearly 1 MeV higher than that of the parent nucleus ($A \approx 200$). Therefore, the binding energy per *nucleon* increases by about 1 MeV, so the kinetic energy released from a single fission is roughly 200 MeV. The magnitude of this release is apparent if one recalls that the energy released in a typical chemical process, such as the burning of coal, is only a few eV for each separate reaction.

Given that the fission of heavy nuclei is so energetically favorable, we must ask why heavy nuclei do not immediately fission into two smaller fragments. To answer this, we must consider the potential energy of the two fragments during the fission process. It is found that this process is quite analogous to the breakup of a charged liquid drop, as is suggested by the sketches in Fig. 17.13. In the first picture the original parent nucleus is shown in its normal, approximately spherical shape. Internal oscillations can cause the nucleus to elongate (second picture) and develop a distinct neck (third picture). At about this stage the Coulomb repulsion of the two pieces begins to dominate the weakening nuclear attraction and flings the pieces apart as two separate nuclei.

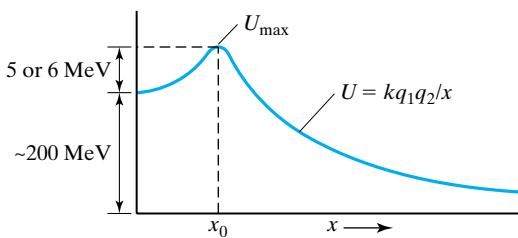
In Fig. 17.14 we show the potential energy of the two fragments as a function of their separation distance x . To understand this curve it helps to imagine the fission process taking place in reverse, starting with the two fragments far apart. When x is very large ($x \rightarrow \infty$) the two fragments exert no force on one another, and we can choose U to be zero. If we now wish to move the two fragments together, we must do work against their Coulomb repulsion, and they gain a potential energy $U = kq_1q_2/x$. When the fragments come within 1 or 2 fm (surface to surface), the attractive nuclear force comes into play and eventually exceeds the Coulomb repulsion. Therefore, the potential energy reaches a maximum at some separation x_0 and then drops again for $x < x_0$, as shown in Fig. 17.14. If the two fragments are closer than x_0 , and if their energy is less than the maximum U_{\max} , they are trapped and fission is impossible. For this reason, the peak in U near $x = x_0$ is called the **fission barrier**; this barrier is the reason that all naturally occurring nuclei are stable against prompt fission.[†]

Neutron-Induced Fission

We see in Fig. 17.14 that 5 or 6 MeV of additional energy is needed to bring a typical heavy nucleus to the top of the fission barrier. Therefore, if one could produce a nucleus in an *excited state* 5 or 6 MeV above the ground state, it would have enough energy to pass over the barrier, and fission could occur,

*The name *fission* is generally reserved for the case that the two fragments have more or less comparable size. Although α decay is qualitatively similar, it is not generally regarded as fission.

[†]If the nucleus were a classical system, then with $E < U_{\max}$, it would be completely trapped inside the fission barrier. As we saw in Section 7.10, quantum mechanics allows the fragments to “tunnel” through this kind of barrier, and this explains the occurrence of spontaneous fission. However, for all the nuclei with which we are concerned here, the probability of spontaneous fission is exceedingly small.



releasing some 200 MeV of kinetic energy. The simplest way to accomplish this uses the capture of slow neutrons, in a process such as



Using the ground-state masses listed in Appendix D, one can easily calculate that the total rest energy of the neutron plus ${}^{235}\text{U}$ is 6.5 MeV greater than that of the ground state of ${}^{236}\text{U}$ (Problem 16.43). Therefore, if the nucleus ${}^{236}\text{U}$ is formed by the process (17.47), it must be formed in an excited state at least 6.5 MeV above its ground state, and in this state the nucleus has enough energy to fission. We can write this sequence of events as



where we have introduced the star to indicate that the ${}^{236}\text{U}$ nucleus is in an excited state.

Since 99% of all natural uranium is the isotope ${}^{238}\text{U}$, it is natural to consider the possibility of neutron-induced fission of ${}^{238}\text{U}$.



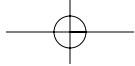
This process is closely analogous to (17.48). Nevertheless, there is an important difference: Recall that the pairing effect means that the ground states of even-even nuclei are somewhat better bound than those of even-odd nuclei. Now, in the process (17.48) the initial nucleus ${}^{235}\text{U}$ is even-odd, while the final nucleus ${}^{236}\text{U}$ is even-even. Thus, because of the pairing effect, the final nucleus in this process has a lower ground state than it otherwise would, and (for a given neutron energy) the excitation energy of the ${}^{236}\text{U}^*$ is therefore higher than it otherwise would be. In the process (17.49), things are the other way round (initial nucleus even-even, and final nucleus even-odd); in this case, the excitation energy of the final ${}^{239}\text{U}^*$ is somewhat lower than it would otherwise be. The upshot is that even at low neutron energies, the nucleus in (17.48) has enough excitation energy to fission, whereas that in (17.49) does not. [And even at higher neutron energies, the probability of fission occurring in (17.49) turns out to be very small.]

This difference between the processes (17.48) and (17.49) means that ${}^{238}\text{U}$ is usually not suitable as a fuel for nuclear fission. It was for this reason that the production of a uranium bomb in World War II required the separation of the rare isotope ${}^{235}\text{U}$. For the same reason, many nuclear reactors that generate electricity require, for their fuel, uranium in which the proportion of ${}^{235}\text{U}$ is at least partially enriched.

The probability for the neutron-induced fission (17.48) of ${}^{235}\text{U}$ is sufficiently large that Fermi undoubtedly produced fission when he bombarded

FIGURE 17.14

The fission barrier (not to scale). The potential energy U of two fission fragments in a typical fissionable nucleus such as ${}^{236}\text{U}$ is shown as a function of the separation distance x . For x large, the repulsive Coulomb force dominates, and $U = kq_1q_2/x$. For small separations, the attractive nuclear force dominates and U forms a potential well. The maximum of U at separation x_0 is called the fission barrier, and the ground state of the parent nucleus is trapped inside this barrier with energy $E < U_{\max}$.



62 Chapter 17 • Radioactivity and Nuclear Reactions

Meitner Bio MANCA ORIGINALE



MANCA ORIGINALE

uranium with neutrons in 1934, but he failed to identify the process. In fact, the possibility of fission was not recognized for another five years even though neutron bombardment of uranium was studied intensively. This surprising delay had two causes: First, all known nuclear reactions produced changes of Z by one or two units at most; thus neutron bombardment of uranium (with $Z = 92$) was expected to produce elements in the neighborhood of $Z = 92$, but certainly not in the range $Z = 35$ to 60. Second, it turns out that many of the common fission fragments are chemically similar to the neighbors of uranium. For example, two typical fragments are strontium ($Z = 38$) and barium ($Z = 56$); a glance at the periodic table shows that these are in the same group as, and hence easily mistaken for, radium ($Z = 88$). Given the small quantities involved, it is easy to see why these fission fragments were wrongly identified as being neighbors of uranium.

The credit for discovering fission is generally given to the Germans Otto Hahn and Fritz Strassman, who established (in 1938) the presence of the fission product barium after bombarding uranium with neutrons. However, the correct interpretation of their discovery (and the name fission) came from the Austrians Lise Meitner and her nephew Otto Frisch in 1939. It was then quickly recognized that fission could possibly be induced on a large enough scale to produce macroscopic amounts of energy, either for explosives or for the generation of electricity.

Chain Reactions

The possibility of large-scale fission depends on two circumstances: The first is that fission can be induced by neutrons, as in (17.48); the second is that the fission reaction itself produces additional neutrons. Once the reaction begins, it can therefore sustain itself in what has come to be called a **chain reaction**, the neutrons produced in each fission going on to induce further fissions. This possibility was recognized almost as soon as fission was discovered,* and by 1942 a group led by Fermi in Chicago had constructed the first nuclear chain reactor, using nearly 50 tons of uranium and uranium oxides.

To see why fission tends to produce neutrons, we consider, yet again, the valley of stability in a plot of Z against N (Fig. 17.15). Let us suppose for a moment that the parent divides into just two fragments, for example,



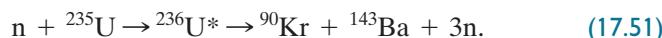
Since the protons and neutrons are uniformly distributed in a heavy nucleus, the proton-to-neutron ratio will be approximately the same for either offspring as for the original parent.

$$(Z/N)_{\text{offspring}} = (Z/N)_{\text{parent}}$$

This means that the two offspring lie close to the line $Z/N = \text{constant}$ shown in Fig. 17.15. Because of the way the valley floor curves, both offspring are some distance away from the valley floor on the neutron-rich side.

*Actually the idea of a chain reaction had been predicted in 1934 by the Hungarian physicist Leo Szilard.

Because the fragments have so many excess neutrons, they promptly eject several separate neutrons to bring them closer to the valley floor.* Thus, in place of (17.50), one might observe



The actual number of neutrons produced in a fission can vary. On average, the neutron-induced fission of ${}^{235}\text{U}$ produces 2.5 neutrons, and it is these that make possible further fissions in a chain reaction.

That each fission produces 2.5 neutrons is not by itself enough to guarantee that the reaction will be sustained. The important parameter is the number, k , of neutrons from each fission that do in fact induce another fission. If this number, called the **reproduction factor**, is less than 1 ($k < 1$), the reaction will die out; if k is greater than or equal to 1 ($k \geq 1$), the reaction will continue; and if k is much greater than 1, the reaction will accelerate explosively.

The value of the reproduction factor k depends on several things. For example, the probability of a neutron-induced fission like (17.51) is much larger for slow neutrons ($K \approx 0.01 \text{ eV}$) than for the fast neutrons ($K \approx 1 \text{ MeV}$) that usually emerge from the fission, as can be seen in Fig. 17.16. Therefore, the fuel in reactors is usually embedded in a **moderator**, a material (such as carbon or water) containing light nuclei that carry away a significant fraction of a neutron's energy as they recoil from elastic collisions (Problem 17.51). Repeated collisions slow down the neutron and improve its chance of inducing a fission before it escapes or gets captured. In this way, the moderator increases the reproduction factor k .

Perhaps the most obvious factor on which k depends is the size of the fissionable sample. In a small sample, most neutrons will escape before they have the opportunity to induce a fission; the larger the sample, the greater the chance for a neutron to induce a fission before it can escape. Under given conditions (density and shape of the material, for example) there is therefore a **critical mass**, below which k is less than 1 and a chain reaction dies out, but above which $k \geq 1$ and the chain reaction is sustained. For example, the critical mass for an isolated sphere of pure ${}^{235}\text{U}$ is on the order of 50 kg.

In reactors that are used to generate electricity it is obviously important to control k so that it remains very close to 1. This is usually done with *control rods*, made of a material such as cadmium that absorbs neutrons strongly. When the rods are fully inserted in the reactor, they absorb many neutrons and k is less than 1. The rods can be withdrawn until $k = 1$ and can be continually adjusted to keep k close to 1. As the reaction proceeds, the kinetic energy of the fission fragments is shared among the atoms of the fuel rods, whose temperature rises. The heat is removed by a coolant, such as water (which also acts as a moderator to slow down the neutrons). The coolant is usually sealed inside the reactor system, but transfers the heat to an external steam boiler which drives the turbines of an electrical generating plant.

Most commercial reactors today use ${}^{235}\text{U}$ as their fuel (either in its normal mix with ${}^{238}\text{U}$ or somewhat enriched). Because of the limited supply of this isotope, such reactors cannot possibly meet society's energy needs indefinitely. One alternative is to use plutonium 239. This fissions even more easily

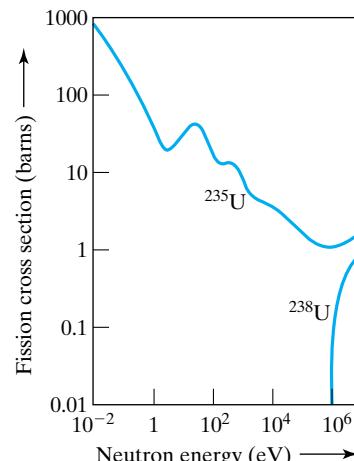
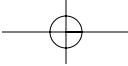


FIGURE 17.16

The cross sections for neutron-induced fission of ${}^{235}\text{U}$ and ${}^{238}\text{U}$. (Recall that the cross section for nuclear reactions is just a convenient measure of their probability.) The cross section for ${}^{235}\text{U}$ rises to very large values at low neutron energies; that for ${}^{238}\text{U}$ is zero below about 1 MeV. Note that both scales are logarithmic. Both curves have many rapid fluctuations that do not show on this scale (and are unimportant in most applications).

*Even so, the offspring still have an excess of neutrons and will undergo several β^- decays. This is why the waste products of fission reactors are extremely β^- radioactive.

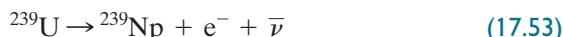


64 Chapter 17 • Radioactivity and Nuclear Reactions

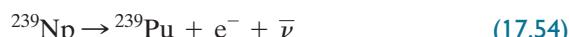
than ^{235}U and, although not naturally available, can be produced from the relatively plentiful ^{238}U by the neutron capture



followed by two β^- decays,



and



A single reactor can be designed in which ^{238}U is converted to ^{239}Pu and ^{239}Pu fissions to produce heat. Some of the neutrons from the fission of ^{239}Pu induce further fissions, while others convert ^{238}U into more ^{239}Pu . Reactors of this type are called **breeder reactors**, and a few are in operation (although not in the United States).

17.8 Nuclear Fusion

We saw at the beginning of Section 17.7 that energy is released when two light nuclei fuse to form a heavier nucleus. An example of such **fusion** is the reaction



in which two hydrogen nuclei (one the ordinary ^1H , and one the heavy isotope ^2H , called deuterium) fuse to form helium 3. Reactions of this kind produce the energy radiated by most stars, including our own sun. Some 50 years ago the same kinds of reaction were first used in the hydrogen bomb, and since then extensive research has been devoted to harnessing fusion reactions for peaceful uses. The principal attraction of fusion as a source of power is that the fuels — light elements like hydrogen — are abundant and cheap. Another important advantage is that it would produce far less radioactivity than fission reactors.

Unfortunately, there is an enormous obstacle to the large-scale use of nuclear fusion. The two initial nuclei are positively charged, and large initial energies (many keV at least) are needed to overcome the Coulomb repulsion and allow fusion. This is easily accomplished on a small scale using an accelerator; but the number of reactions produced in this way is hopelessly insufficient, even to replace the energy needed to run the accelerator.

It seems that the only way to produce macroscopic amounts of energy by nuclear fusion is to heat the fuel to high temperatures, where the kinetic energy of thermal motion is sufficient to overcome the Coulomb barrier. (This is the origin of the name **thermonuclear reaction** — a nuclear reaction induced by high temperature.) Unfortunately, the temperatures required are extremely high, as we can easily see: For two hydrogen nuclei (ordinary or heavy) to react, they must certainly come within 10 fm of one another (and probably closer). At this separation their potential energy is

$$U = \frac{ke^2}{r} = \frac{1.44 \text{ MeV} \cdot \text{fm}}{10 \text{ fm}} = 144 \text{ keV}$$

Thus, to bring them close enough in a head-on collision, we must give each nucleus about 70 keV of kinetic energy. Now, we know from kinetic theory (Section 3.7) that the average energy of a particle at absolute temperature T is

$$\langle K \rangle = \frac{3}{2} k_B T \quad (17.56)$$

where k_B is Boltzmann's constant, $k_B = 8.62 \times 10^{-5}$ eV/K (and K denotes the kelvin, the SI unit of thermodynamic temperature). Thus if the average thermal energy is to be about 70 keV, the required temperature is

$$T = \frac{2}{3} \frac{K_{av}}{k_B} = \frac{2}{3} \times \frac{7 \times 10^4 \text{ eV}}{8.6 \times 10^{-5} \text{ eV/K}} \approx 5 \times 10^8 \text{ K} \quad (17.57)$$

The temperature (17.57) is actually an overestimate of the temperature needed to produce thermonuclear reactions. In the first place, (17.56) is the average kinetic energy at temperature T ; even at lower temperatures, some particles will still have this much energy. Second, because of the quantum effect called tunneling (which we met in Section 7.10 and will take up again in Section 17.10), even particles with insufficient energy to surmount the Coulomb barrier *can* undergo fusion. For both these reasons, fusion can occur at temperatures an order of magnitude below (17.57) at about*

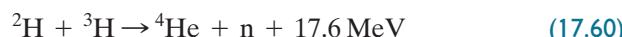
$$T \approx 5 \times 10^7 \text{ K} \quad (17.58)$$

still a very high temperature!

The problems of producing temperatures of order 10^7 K on earth are obvious. The only successes so far are, unfortunately, in nuclear explosions. A fission bomb can produce temperatures of this order and act as the trigger for a fusion bomb. The principal fuel for the fusion bomb is deuterium, or heavy hydrogen, ${}^2\text{H}$ (hence the name "hydrogen bomb"). One of the fusion reactions is the DD reaction



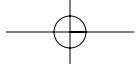
which produces the isotope ${}^3\text{H}$, called tritium. This in turn can fuse with deuterium in the deuterium-tritium (DT) reaction:



Notice that although the energy release in these fusion reactions is smaller than the 200 MeV released in fission, the energy released *per nucleon* (and hence per unit mass) is larger in fusion than in fission.

The production of controlled thermonuclear reactions for the generation of electricity is much more difficult than producing an explosion. The difficulty is to contain the fuels while they are heated to many millions of degrees. Obviously, any conventional container would vaporize long before such temperatures were reached. Two methods of confinement are being investigated: a

*The precise temperature needed naturally depends on the circumstances; for example, the temperature in the sun (where thermonuclear reactions occur rather slowly) is somewhat cooler.



66 Chapter 17 • Radioactivity and Nuclear Reactions

Hans Bethe
(Born 1906, German-American)

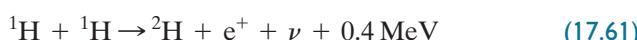


After postdoctoral work with Rutherford in Cambridge and Fermi in Rome, Bethe taught in Germany for a few years before coming to the United States in 1935. Among many contributions to atomic and nuclear physics, he is best known for finding the two nuclear cycles by which most stars get their energy. For this discovery, he won the 1967 Nobel Prize in physics.

confinement by magnetic fields, and a brief compression by bursts from several high-power lasers. After some 50 years of intense research in several countries, it is still not clear that nuclear fusion will be successfully harnessed for the peaceful generation of electricity.

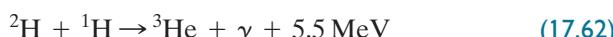
Fusion in Stars

The temperatures needed to produce thermonuclear fusion are available naturally inside stars, and nuclear fusion is indeed the principal source of stellar energy.* A relatively young and small star like our own sun consists mainly of hydrogen, and the main source of energy is a sequence of three reactions called the **proton–proton cycle**, first worked out by Hans Bethe in 1938. The first reaction produces deuterons (^2H nuclei).

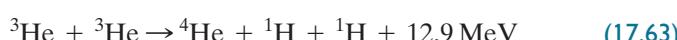


This reaction involves the conversion of a proton to a neutron by β^+ decay, during the fleeting collision of the two protons, and the reaction (17.61) is therefore extremely unlikely (Problem 17.54). Nevertheless, proton–proton collisions are so frequent in the star that this reaction proceeds at a sufficient rate.

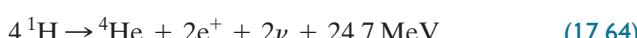
A deuteron produced in the reaction (17.61) can next collide with another proton to produce helium 3.



Finally, two of the ${}^3\text{He}$ nuclei produced in this way can collide to give the very stable ${}^4\text{He}$.



Notice that the net effect of the whole proton–proton cycle is to convert four protons into a ${}^4\text{He}$ nucleus (plus assorted electrons and neutrinos)



(The two positrons quickly annihilate with two electrons to produce a further 2.0 MeV, so the total yield is actually more like 27 MeV.) In this way, the sun generates about 4×10^{20} megawatts of power and converts some 50 trillion tons of hydrogen into helium every day.[†]

When a star has “burned” most of its hydrogen to helium, its source of energy is temporarily exhausted, and the star can no longer support itself against its own gravitational attraction. The star therefore contracts, and its central temperature rises until new fusion processes, forming carbon from helium, can begin. When the helium is exhausted in large stars, there is a second transition to processes in which carbon is fused to form heavy elements all the way up to iron. Since iron has the highest value of B/A , it is the heaviest

*Before nuclear reactions were understood, the source of stellar energy was one of the outstanding mysteries of astronomy. For example, it is easy to show that if stars depended on gravitational potential energy, they would collapse in just a few million years, whereas we know that our own sun is several billion years old.

[†]In heavier stars the conversion of hydrogen to helium is accomplished primarily by a different sequence of reactions, called the carbon cycle (Problem 17.53).

element that can be formed by fusion with the release of energy.* Heavier elements, from iron to uranium and beyond, are formed by a succession of neutron captures followed by β decays. This process can occur slowly in an old but stable star or very rapidly in a supernova explosion. This type of explosion occurs when a heavy star, having exhausted its supply of fusion fuel, undergoes a catastrophic collapse and then explodes. The explosion ejects an enormous outer shell of matter and leaves behind a tiny but massive neutron star or black hole. The matter that is ejected, including some of the recently formed heavier elements, is dispersed through interstellar space and, in due course, is formed into new stars like our own sun. The presence of the heavy elements in the sun and planets is evidence that they, and we, are descendants of many long-dead stars.

17.9 Nuclear Weapons

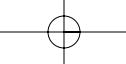
Fission Weapons

The preceding sections on nuclear fission and fusion provide a background for a discussion of nuclear weapons. Fission weapons were developed during World War II soon after the discovery of fission and the chain reaction. When it became clear that fission could release a million times more energy than conventional high explosives, fission was recognized as a potential super weapon that could conceivably determine the outcome of the war. Scientists of several nations were so concerned about the potential power of such weapons that many stopped open publication of their research papers. The United States started an intensive and secret research program to develop an atomic bomb. Efforts were also started in Great Britain, Germany, the Soviet Union, and Japan, but the United States placed more resources into its program, motivated by a fear that other nations would develop “the bomb” first.

The basic concept of the weapon was the assembly or compression of subcritical masses of U-235 or Pu-239 to create a supercritical mass that could sustain an exponentially growing chain reaction and produce a nuclear explosion. At first glance it would seem that a simple machine could push the subcritical masses together. However, as the masses get closer together, stray neutrons from cosmic radiation or spontaneous fission can trigger chain reactions at the wrong time for optimum yield. Worse yet, if the assembly is too slow, there will be time for the masses to heat and either warp or melt, thereby interrupting the formation of a chain reaction. To prevent the occurrence of such a nuclear fizzle, the subcritical masses must be assembled very quickly (a few microseconds). This is accomplished by using explosives to assemble the weapon.

In the **gun-type** weapon [Fig. 17.17(a)], one of the masses is fired from a gun barrel and fits like a plug into the other mass, so that the combined pair forms a critical mass. The gun design does not work with plutonium because the spontaneous fission rate for Pu is so high that subcritical Pu masses cannot be fired together fast enough to avoid a fizzle. Instead, an **implosion method** is used, in which chemical explosives (TNT) surround a single subcritical sphere of plutonium [Fig. 17.17(b)]. When the explosives are triggered, the in-going

*This is a slight oversimplification. Nickel 62 actually has a slightly higher value for B/A . However, there is no sequence of fusions that leads to ^{62}Ni with any appreciable probability, and ^{56}Fe is the end of the line in normal stars.



68 Chapter 17 • Radioactivity and Nuclear Reactions

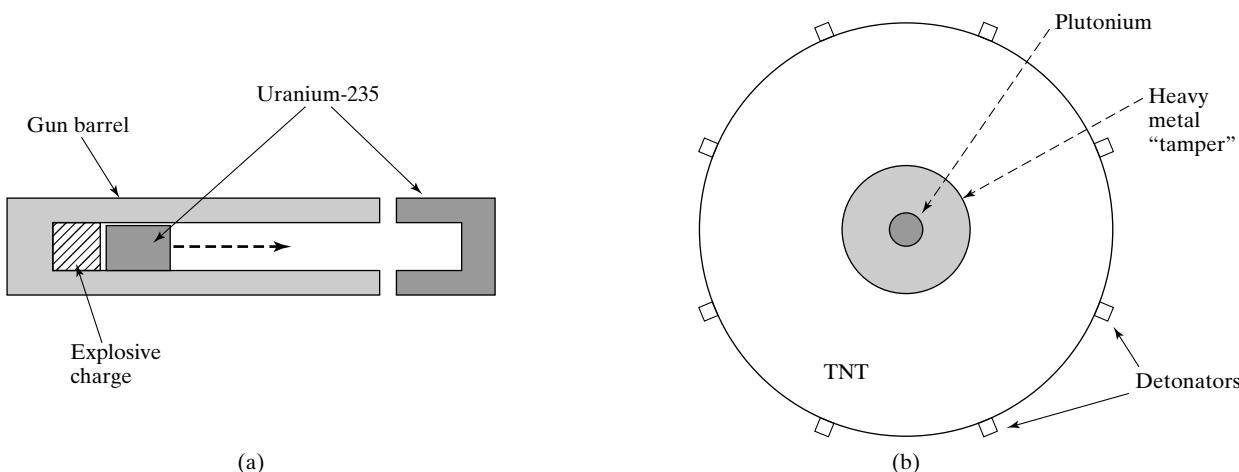


FIGURE 17.17

(a) In the gun-type of fission bomb the subcritical mass on the left is fired into another on the right to form a single supercritical mass. (b) In the implosion type, chemical explosives surrounding the subcritical mass of plutonium are fired simultaneously to compress the plutonium, rendering it supercritical. A thick shell of heavy metal, such as ^{238}U , surrounds the plutonium core and acts as a “tamper,” a massive container that prevents the plutonium from blowing apart too early after the fission process begins.

shock wave with a pressure of a million pounds per square inch compresses the core (about the size of a grapefruit) to twice its original density — more than enough to form a critical mass. Photographs of the original gun device, known as “Little Boy,” and implosion device, called “Fat Man,” are shown in Figure 17.18.

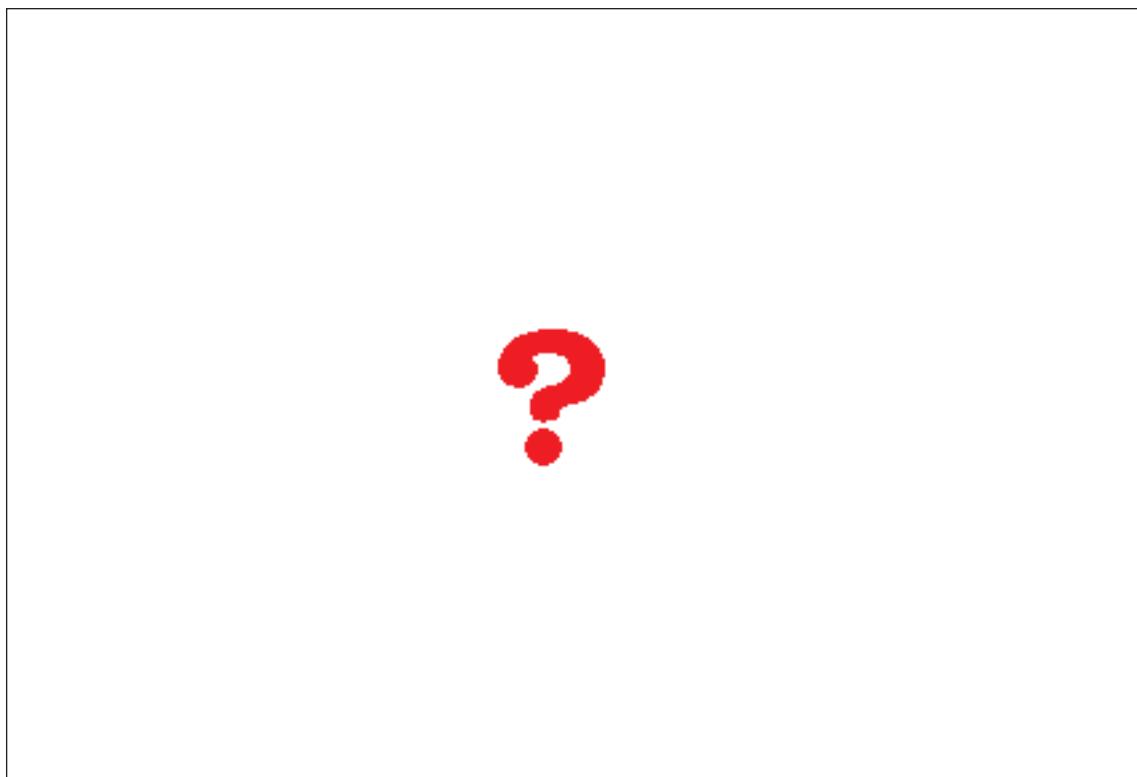
The first test of a nuclear bomb was made in the New Mexico desert on 16 July, 1945, using a prototype of the Fat Man device. The first nuclear weapon used in warfare was Little Boy, which was dropped on Hiroshima on 6 August, 1945. Just three days later, Fat Man was dropped on Nagasaki, bringing World War II to an abrupt end.

The **yield** of a nuclear weapon is the amount of explosive energy released by the weapon. This is generally expressed as the quantity of chemical explosive that would release the same amount of energy. The yields of fission bombs are usually expressed in kilotons of TNT. (1 kiloton of TNT = 4.3×10^{12} J.) The yields of larger fusion devices (hydrogen bombs) are usually expressed in megatons of TNT. The yield of Little Boy was between 12 and 15 kilotons, while that of Fat Man was about 22 kilotons.

Fusion Weapons

Fusion bombs (also called hydrogen, or thermonuclear, bombs) get their enormous destructive power from the fusion of light nuclei. Their principal fuels are the two hydrogen isotopes, deuterium (D) and tritium (T). These isotopes combine in the DD reaction (17.59) and the DT reaction (17.60). As we saw in Section 17.8, the colliding nuclei must have enough energy to overcome their mutual Coulomb repulsion before they can fuse. This requires that the fuel be raised to a temperature between 10 and 100 million kelvins, which is accomplished by using a fission bomb as a “trigger.”

A large explosion also requires that the density of the fuel be higher than is possible with hydrogen gas. The first successful explosion involved building

**FIGURE 17.18**

The three atomic bombs of World War II. (a) The “gadget,” an implosion-type plutonium bomb tested in the New Mexico desert on July 16, 1945. (b) The gun type, called “Little Boy,” dropped on Hiroshima on August 6. (c) The implosion type, called “Fat Man,” dropped on Nagasaki on August 9. The names Little Boy and Fat Man reflect their shapes, which in turn reflect their different functioning as indicated in Fig. 17.17. The Little Boy was 10 feet long and the Fat Man 5 feet in diameter. Both bombs weighed about 5 tons.

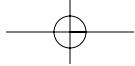
AU:
“gun type”
with hyphen?

a huge refrigerator plant to liquify the fuel. While the resulting explosion yielded more than 10 megatons, it did not lead to a deliverable bomb since no plane could carry around the needed refrigeration. A smaller bomb was eventually made by solidifying the deuterium into lithium-deuteride (the solid formed of ${}^6\text{Li}$ and deuterium). This had the additional advantage that, when bombarded by the neutrons from the fission trigger, the ${}^6\text{Li}$ dissociates into tritium and ${}^4\text{He}$, thus furnishing tritium as a fuel and saving the trouble of making and storing radioactive tritium.

17.10 The Theory of Alpha Decay*

*This section could be omitted without loss of continuity.

As we saw in Section 17.5, the naturally occurring radioactive series involve combinations of α and β decays. All three series originate in the α decays of their parent nuclei ${}^{232}\text{Th}$, ${}^{235}\text{U}$, and ${}^{238}\text{U}$. These nuclei exist on earth today only because they have half-lives of billions of years, so that appreciable amounts have survived the billions of years since the formation of our solar system. These extremely long half-lives pose a puzzle: How can a process driven by nuclear energies of millions of electron volts take billions of years to occur?



70 Chapter 17 • Radioactivity and Nuclear Reactions

TABLE 17.1

Five alpha-emitting nuclei in order of decreasing half-life. The second column shows the kinetic energy released in the decay, and the third shows the half-life.

Nucleus	K(MeV)	$t_{1/2}$
^{232}Th	4.1	14 billion yr
^{226}Ra	4.9	1600 yr
^{240}Cm	6.4	27 days
^{194}Po	7.0	0.7 s
^{216}Ra	9.5	0.18 μs

The long half-lives of ^{232}Th , ^{235}U , and ^{238}U are not the only puzzles posed by α decay. Table 17.1 shows the half-lives of 5 different α -radioactive nuclei. You can see that the half-lives vary over an astonishing range, from billions of years to fractions of a microsecond, while the energies released range only from about 4 to about 9 MeV. As you can also see, the half-lives are correlated with the energy released: the shorter the half-life, the higher the energy released and vice versa.

Given these facts, the questions we must answer are these: How do nuclei emit alpha particles? Given that they *can* emit alpha particles, why do some nuclei wait more than 10 billion years before doing so? And why is the half-life so extraordinarily sensitive to the energy release K , with a factor of 2 difference in K corresponding to nearly 25 orders of magnitude in $t_{1/2}$? The answers to these questions were found in 1928 by the Russian George Gamow and independently by the Americans Condon and Gurney.

The theory of alpha decay starts from the plausible assumption that two of the protons inside a heavy nucleus can occasionally cluster together with two of the neutrons to form an alpha particle. Experiments* show that this is correct: Alpha particles are continually forming and dissolving inside a nucleus. Averaged over time, it is found that there is of order one alpha particle inside any heavy nucleus.

To find how quickly alpha particles escape from the nucleus, we must first estimate how frequently they appear at the nuclear surface. An alpha particle is formed from four nucleons, each with potential energy about -50 MeV and kinetic energy about 25 MeV. Therefore, the alpha particle should have a kinetic energy of roughly 100 MeV, corresponding to a speed $v \sim c/4$. For a nuclear diameter D of about 15 fm, the frequency with which an alpha particle appears at the surface of the nucleus would be roughly

$$f \sim \frac{v}{D} \sim \frac{c/4}{15 \text{ fm}} \sim 5 \times 10^{21} \text{ s}^{-1} \quad (17.65)$$

We see that alpha particles appear at the nuclear surface more than 10^{21} times per second. We would not necessarily expect these particles to escape since the strong nuclear force pulls them back into the nucleus. Nevertheless, the observed fact is that an alpha particle does occasionally escape. Our task is to see how this occurs and to find the probability of its doing so.

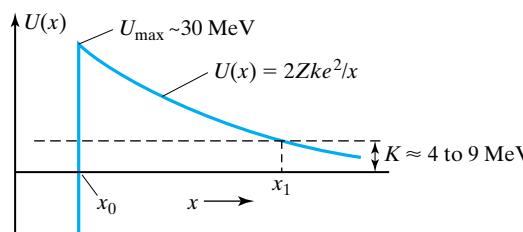
Let us denote by P the probability that a single alpha particle striking the nuclear surface will escape. Our ultimate goal is to find the decay constant, r ,

*For example, reactions involving ejection or transfer of alpha particles from nuclei.

George Gamow
(1904–1968, Russian-American)



After various postdoctoral positions in Europe, Gamow settled in the United States. In addition to his theory of alpha decay, he worked out many details of stellar evolution, was an early proponent of the “big-bang” origin of the universe, and was the first to guess (correctly) that the genetic code is carried by triplets of nucleotides. Gamow was also the author of many entertaining books on science.



of the nucleus. Since r is the probability per unit time of an alpha particle escaping, it is the product of f (the frequency with which alpha particles collide with the surface) and P (the probability of escape in any one collision).

$$r = fP \quad (17.66)$$

We have already estimated f in (17.65), and we must now find the probability P .

As we know, the probability of finding a quantum particle in some region is given by the square of its wave function, and the wave function is determined by the Schrödinger equation, whose solution requires knowledge of the potential-energy function. The potential energy of the alpha particle is sketched in Fig. 17.19 and is qualitatively similar to the potential energy of two fission fragments (shown in Fig. 17.14). If the alpha particle is outside the nucleus, then U is just the Coulomb potential energy $U(x) = 2Zke^2/x$, where x is the separation of the alpha from the residual nucleus and Ze is the charge of the residual nucleus. (Remember, the charge of the alpha is $2e$.) For x less than some x_0 , about equal to the nuclear radius R , the attractive nuclear force dominates, and the potential energy drops abruptly. There is therefore a *barrier*, exactly analogous to the fission barrier, and it is this barrier that prevents the alpha particle from escaping instantly. However, as we saw in Section 7.10, quantum mechanics allows a particle to *tunnel* through a barrier of this type, and this quantum tunneling is what allows the alpha particle to pass through the barrier and escape.

To calculate the probability of such tunneling, we must establish the approximate parameters of the barrier confining the alpha particle. Its left-hand boundary is at the point shown as x_0 , which is about equal to the nuclear radius (or, more precisely, the sum of the radii of the residual nucleus and the alpha).

$$x_0 \approx 8 \text{ fm}$$

As can be seen in Fig. 17.19, the height U_{\max} of the barrier is simply the electrostatic potential energy at $x = x_0$:

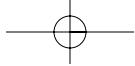
$$U_{\max} = 2Z \frac{ke^2}{x_0} \sim 2 \times 90 \times \frac{1.44 \text{ MeV} \cdot \text{fm}}{8 \text{ fm}} \approx 30 \text{ MeV}$$

We know that when the alpha particle emerges from the nucleus and moves far away, it has kinetic energy and hence total energy between 4 and 9 MeV.* By conservation of energy, this must be the same as its energy before the decay occurred. Therefore, its energy inside the nucleus was also between 4 and 9 MeV (as shown by the dashed horizontal line in Fig. 17.19). This is far

FIGURE 17.19

The barrier that keeps an alpha particle from escaping promptly from a nucleus. The curve shown is the potential energy $U(x)$ of the alpha particle at a distance x from the center of the offspring nucleus. For $x > x_0$, it is just the Coulomb potential, while for $x < x_0$, it is dominated by the attractive nuclear force and is strongly negative. (The actual curve would be somewhat rounded at x_0 .) The observed kinetic energy when the alpha particle escapes is shown as K .

*We are treating the process nonrelativistically, as is certainly an excellent approximation since $mc^2 \approx 4000 \text{ MeV}$ for an alpha particle. Thus “total energy” here means the classical energy, $E = K + U$, and if $U = 0$, then $E = K$.



72 Chapter 17 • Radioactivity and Nuclear Reactions

below the top of the barrier ($U_{\max} \approx 30 \text{ MeV}$) — which confirms that if the alpha were a classical particle inside the nucleus, it would be permanently trapped by the barrier.

Another important parameter shown in Fig. 17.19 is the distance x_1 . This is the separation at which an alpha particle coming *inward* with initial kinetic energy K would be stopped by the barrier. The value of x_1 is determined by the condition

$$2Z \frac{ke^2}{x_1} = K$$

or

$$x_1 = 2Z \frac{ke^2}{K} \approx 2 \times 90 \times \frac{1.44 \text{ MeV} \cdot \text{fm}}{4 \text{ to } 9 \text{ MeV}} \approx 30 \text{ to } 65 \text{ fm}$$

The interval from x_0 to x_1 is the classically forbidden region. Its length, $L = x_1 - x_0$, is the thickness of the barrier through which the alpha particle must tunnel. By nuclear standards, L is rather long and varies appreciably (from roughly 20 to 55 fm as K ranges from 9 down to 4 MeV). We will see that these are the main reasons why the escape probability P is so low and why it varies so much with the energy of the emerging alpha particle.

Barrier Penetration

In Section 7.10 we saw that a quantum particle can pass through a potential-energy barrier that would completely block a classical particle. For a simple rectangular barrier, as shown in Fig. 17.20, we found the probability that a particle which strikes the rectangular barrier of Fig. 17.20 with energy $E < U_0$ will tunnel through and emerge on the outside is

$$P \approx e^{-2\alpha L} \quad (17.67)$$

where

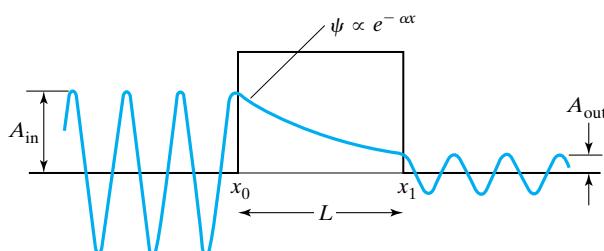
$$\alpha = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}} \quad (17.68)$$

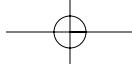
and m is the mass of the particle concerned.

The escape probability (17.67) depends on two variables: α , as given by (17.68), and L , the barrier thickness. If the barrier is high and wide (as is the case in alpha decay), both α and L are large and the escape probability is very small. (Remember that if x is large, e^{-x} is extremely small; for example, if $x = 10$, then $e^{-x} = 0.00005$.) Furthermore, if α is large, $P = e^{-2\alpha L}$ is very

FIGURE 17.20

Wave function for a barrier of finite length. Inside the well ($x < x_0$), $\psi(x)$ is sinusoidal, with amplitude A_{in} ; inside the barrier, it decreases exponentially; then for $x > x_1$, it is sinusoidal again, with amplitude A_{out} .



Section 17.10 • The Theory of Alpha Decay **73**

sensitive to the barrier thickness L . This is the main reason that the observed half-lives vary over such an enormous range.

Equation (17.67) gives the escape probability for any particle confined by a rectangular barrier. Our interest is in an alpha particle confined by the nonrectangular barrier of Fig. 17.19. To find the corresponding escape probability, we can approximate the actual barrier by a succession of n rectangular barriers as in Fig. 17.21. The total escape probability is the product of the n individual probabilities

$$\begin{aligned} P &= P_1 P_2 \cdots P_n \\ &= e^{-2\alpha_1 \Delta x_1} e^{-2\alpha_2 \Delta x_2} \cdots e^{-2\alpha_n \Delta x_n} = \exp\left(-2 \sum \alpha_i \Delta x_i\right) \quad (17.69) \end{aligned}$$

where each α_i is given by (17.68) with the appropriate barrier height U_i . In the limit that all $\Delta x_i \rightarrow 0$, the sum in (17.69) becomes an integral, and we obtain the desired probability,

$$P = \exp\left[-2 \int_{x_0}^{x_1} \alpha(x) dx\right] \quad (17.70)$$

where $\alpha(x)$ is found from (17.68),

$$\alpha(x) = \sqrt{\frac{2m[U(x) - K]}{\hbar^2}} \quad (17.71)$$

by replacing U_0 with $U(x)$ and noting that E is just K , the kinetic energy when the alpha particle is far away from the nucleus and U is zero.

For the case of alpha decay, we know that $U(x) = 2Zke^2/x$, and the integral in (17.70) can be evaluated (Problem 17.60) to give

$$P = \exp\left[-\frac{2\pi ke^2\sqrt{2m}}{\hbar} ZK^{-1/2} + \frac{8\sqrt{mke^2}}{\hbar} (ZR)^{1/2}\right] \quad (17.72)$$

where Z and R denote the charge and radius of the residual nucleus. It is convenient to rewrite this result as

$$P = \exp[-aZK^{-1/2} + b(ZR)^{1/2}] \quad (17.73)$$

where the two constants a and b are easily evaluated (Problem 17.58) to give

$$a = \frac{2\pi ke^2\sqrt{2m}}{\hbar} = 3.97 \text{ (MeV)}^{1/2} \quad (17.74)$$

and

$$b = \frac{8\sqrt{mke^2}}{\hbar} = 2.98 \text{ (fm)}^{-1/2} \quad (17.75)$$

If we substitute the typical values $Z = 90$, $K = 6 \text{ MeV}$, and $R = 8 \text{ fm}$ into (17.73), we find that

$$P \approx \exp(-146 + 80) \approx 2 \times 10^{-29}$$

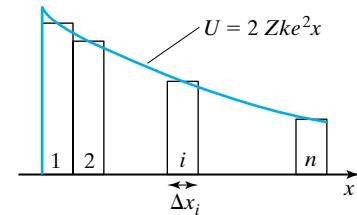
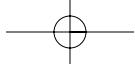


FIGURE 17.21

The actual curved barrier can be approximated as n rectangular barriers, each with thickness Δx_i .



74 Chapter 17 • Radioactivity and Nuclear Reactions

As anticipated, the probability P that an alpha particle striking the surface of a nucleus will escape is extremely small.

Half-Lives

Since experimental data are generally given in terms of half-lives, let us assemble our results to give an expression for the half-life of an alpha-emitting nucleus. The half-life is $(\ln 2)/r$ and the decay constant r is $r = fP$, where f is the frequency with which alpha particles strike the nuclear surface. Therefore,

$$t_{1/2} = \frac{\ln 2}{r} = \frac{0.693}{fP}$$

Since P is given by (17.73) as an exponential, it is convenient to take natural logarithms to give

$$\ln t_{1/2} = -\ln P + \ln \frac{0.693}{f}$$

or, from (17.73),

$$\ln t_{1/2} = aZK^{-1/2} - b(ZR)^{1/2} + c \quad (17.76)$$

In this formula Z and R are the charge and radius of the residual nucleus. The constants a and b are given by (17.74) and (17.75). The constant c is $\ln(0.693/f)$, where f was estimated as roughly $5 \times 10^{21} \text{ s}^{-1}$ [see (17.65)]; therefore,

$$c = \ln \frac{0.693}{f} \approx -50 \quad (17.77)$$

The simplest way to test our theory is to consider a set of alpha-emitting nuclei that all have the same value of Z . For example, we could consider the isotopes ^{192}Po , ^{194}Po , ^{196}Po , ^{208}Po , ^{210}Po , ^{212}Po , ^{214}Po , ^{216}Po of polonium, all of which decay exclusively by alpha emission.* Since these nuclei all have the same value of Z and very nearly the same radius R , the second term in (17.76) is approximately constant, and (17.76) reduces to

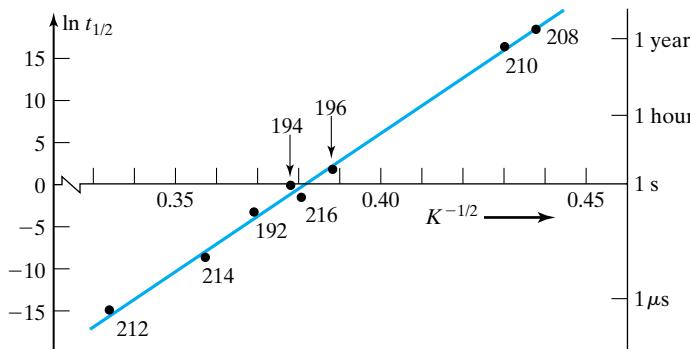
$$\ln t_{1/2} = aZK^{-1/2} + \text{constant} \quad (17.78)$$

That is, our theory of alpha decay predicts that a plot of $\ln t_{1/2}$ against $K^{-1/2}$ for a set of isotopes should be a straight line. The remarkable success of this prediction can be seen in Fig. 17.22, which shows the data for the eight isotopes of polonium mentioned before. The line shown is the least-squares fit to the data. Its slope is 328, in good agreement with the value predicted by (17.78).

$$\text{slope} = aZ = 3.97 \times 82 = 326$$

*We must exclude any isotope with other decay modes in addition to alpha emission, since these reduce $t_{1/2}$. We have also excluded isotopes with A odd since these all have nonzero angular momentum, which (it turns out) complicates the analysis and alters $t_{1/2}$ appreciably.

Section 17.10 • The Theory of Alpha Decay 75

**FIGURE 17.22**

Plot of $\ln t_{1/2}$ against $K^{-1/2}$ for eight alpha-emitting isotopes of polonium. (Half-lives, $t_{1/2}$, in seconds and energy release, K , in MeV.) The number beside each point is the mass number A of the isotope. The line is the least-squares fit to the data, and the axis on the right shows $t_{1/2}$ itself.

The line's intercept with the vertical axis (not shown in the picture) is at -125 , while our predicted value from (17.78) is

$$\text{"constant"} = -b(ZR)^{1/2} + c \approx -120 \quad (17.79)$$

(if we take $R \approx 7$ fm). Given the rough approximations that went into this value, this agreement is entirely satisfactory. The agreement between theory and experiment is even better if we allow for the small variations of R in the "constant" of (17.79).

If we wish to use the result (17.76) to compute half-lives of various different elements, we must recognize that the three variables K , Z , and R can all vary. However, neither Z nor R varies very much. (For example, $81 \leq Z \leq 90$ for all residual nuclei in the natural radioactive series.) Thus, the single most important effect is still the energy dependence of the term $aZK^{-1/2}$ in $\ln t_{1/2}$. Small differences in K correspond to appreciable differences in this term and hence enormous differences in $t_{1/2}$. In particular, as K ranges from 4 to 9 MeV, this term is the main reason that $t_{1/2}$ drops by some 24 orders of magnitude, as noted at the beginning of this section.

Spontaneous Fission

We have already noted the similarity of alpha decay and nuclear fission. In particular, for both processes, there is a barrier that prevents the decay from occurring instantaneously. In Section 17.7 we discussed *induced fission*, in which incident neutrons produce a nucleus with sufficient excitation energy to pass immediately over the top of the fission barrier. It should now be clear that even without any excitation energy, nuclei should undergo **spontaneous fission** when the fission fragments tunnel through the fission barrier.

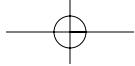
The probability of spontaneous fission can be calculated using the theory of barrier penetration. If we suppose, for simplicity, that the barriers are rectangular, the escape probability is given by (17.67) as

$$P \approx e^{-2\alpha L} \quad (17.80)$$

where

$$\alpha = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}} \quad (17.81)$$

Comparing the fission barrier of Fig. 17.14 and the alpha barrier of Fig. 17.19, we see that the fission barrier is somewhat lower ($U_0 - E$ is less) and shorter



76 Chapter 17 • Radioactivity and Nuclear Reactions

(L is less). Both of these observations would suggest that spontaneous fission would occur more easily than alpha decay. However, there is a third factor working the other way. In (17.81) we see that the parameter α depends on the mass of the emitted particle. Since fission fragments are much heavier than alpha particles, the parameter α is appreciably *larger* for fission than for alpha decay. In all naturally occurring nuclei this effect is dominant, and spontaneous fission is much less probable than alpha decay. Nevertheless, spontaneous fission has been observed in several nuclei, and is actually the dominant decay mode in some artificial nuclei, such as ^{250}Cm and ^{258}No .

Other examples of tunneling occur in solid-state physics. For example, it plays an essential role in the scanning tunneling microscope, or STM, which was described in Section 14.7.

17.11 Accelerators*

* If you plan to read Section 18.11 on high-energy accelerators, you should definitely read this section. Otherwise, this section could be omitted without loss of continuity.

By far the most fruitful source of information about nuclei is the *scattering*, or *collision experiment*, in which a beam of energetic projectiles, such as alpha particles or protons, is directed at a target containing the nuclei to be studied. The projectiles may simply scatter off the target, or they may induce various reactions; but in either case one can infer many properties of the nucleus by studying the results of the collisions. Almost all the great discoveries described in this chapter — of the nucleus, of the neutron, of fission, and many more — were collision experiments of this kind. In this section we describe **accelerators**, the machines that accelerate the projectiles to the energies required for interesting collisions. In Section 17.12 we describe the detectors that allow us to identify the particles produced in collisions and to measure their properties.

The earliest collision experiments all used alpha particles produced by naturally occurring radioactive nuclei. These alpha particles had energies restricted to a few values in the range from 4 to 9 MeV. By the 1920s it was clear that it would be useful to do collision experiments using particles whose energy could be varied continuously over a larger range, and with other projectiles besides alpha particles. It was for this reason that several physicists set out to design “accelerators” to accelerate a variety of particles to various energies. Since then many different accelerators have been developed, of which we describe a few. All these machines use electric fields to accelerate the projectiles, which must therefore be electrically charged. The projectiles used include protons, electrons, and the positive and negative ions of many different atoms. The general principles are the same whether the particle’s charge is positive or negative; to be definite, we will usually suppose in what follows that the particle is positive, that is, is a proton or other positive ion.

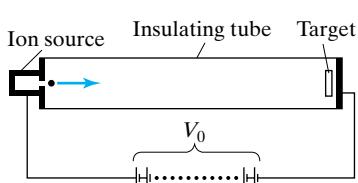


FIGURE 17.23

In a DC accelerator a large potential difference V_0 applied between the ends of an evacuated tube accelerates positive ions from the ion source to the target. If the ion has charge q , this gives it an energy V_0q .

DC Accelerators

The most obvious way to accelerate a charged particle is to use an electric field applied along the length of a single evacuated tube, as sketched in Fig. 17.23. The field is produced by applying a potential difference between the ends of the tube, and since this potential difference is of a fixed sign, we refer to accelerators of this kind as **DC accelerators**.

While DC accelerators are very simple in principle, there was nevertheless a large practical difficulty in their initial development: To find out anything

about nuclear structure, the projectile must approach the target nucleus within 1 or 2 fm (and preferably closer). Because of the Coulomb repulsion between the projectile and the nucleus, this requires a large energy; for example, to bring two protons within 1 fm requires nearly 1.5 MeV. Thus, the accelerator must give the projectile a kinetic energy of order 1 MeV — or at least several hundred keV — to be of any use.* If the projectile has charge e (as is usually the case) and the accelerating potential is V_0 , the kinetic energy given to the projectile is $K = V_0e$. This means that V_0 needs to be at least several hundred kilovolts, and preferably a megavolt or more. In the 1920s there was no known way to produce such large DC voltages. Thus before a useful accelerator could be built, a way had to be found to generate these huge voltages.

In the quest for high voltages, one group went so far as to stretch a cable between two mountaintops to exploit the high potentials found in thunderstorms. A spherical terminal was suspended from the cable, and sparks several hundred feet long jumped from the terminal to the valley floor during thunderstorms. Plans were made to install an accelerating tube but were abandoned when one of the investigators was killed by lightning.

The first successful high-voltage accelerator was produced by the English physicists Cockcroft and Walton, who used AC transformers and rectifiers to put a steady voltage across a bank of capacitors. In this way they accelerated protons to 500 keV and, in 1932, observed the first purely man-made nuclear reaction, described in Section 17.6. Their accelerator is shown in Fig. 17.24. Modern versions of the Cockcroft–Walton accelerator can achieve about 1 million volts and are widely used to inject protons into higher-energy accelerators such as the linear accelerators described later.

Another successful source of high voltages was the electrostatic generator designed by the American physicist Robert van de Graaff in the early 1930s. In this machine the high-voltage terminal is a large hollow conductor, usually spherical or oval in shape. A motor-driven belt carries electrostatic charges to the inside of this terminal, where they are deposited and immediately move to the outside surface. In this way a large total charge can be placed on the terminal and produces a correspondingly high voltage. An early version of an accelerator using a van de Graaff generator is shown in Fig. 17.25. Modern versions enclose the entire device in a tank of high-pressure gas to inhibit loss of charge by sparking; in this way they can achieve voltages of about 10 MV and are used to accelerate many different particles, ranging from electrons to heavy ions such as uranium.

AC Accelerators

A DC accelerator accelerates particles to high energy in a single large boost, produced by a correspondingly large voltage. An obvious alternative is to use a succession of smaller boosts. This removes the need for such enormous voltages, and if one uses *many* small boosts, a higher final energy can be achieved than is possible with any DC accelerator. Most accelerators in use today are of this multiple-boost type. Since they all use an alternating current to produce

*We are assuming here that the projectile is positive and is therefore *repelled* by the target nucleus. One might think that one could get around this large energy requirement by using a negative projectile, but this is not so: The only possibility is an electron, and because of its low mass, a low-energy electron has a wavelength far bigger than a nucleus. This makes low-energy electrons almost useless as probes of nuclei. In fact, an electron needs an energy of hundreds of MeV before its wavelength is comparable to the nuclear size.



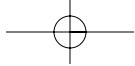
FIGURE 17.24

The accelerating tube of Cockcroft and Walton's proton accelerator can be seen above the small enclosure where the alpha particles from the reaction $p + {}^7\text{Li} \rightarrow \alpha + \alpha$ were observed on a scintillating screen. The observer in this photograph is Cockcroft himself.



FIGURE 17.25

An early van de Graaff accelerator. The belt that carried charges into the egg-shaped terminal is visible at the upper right; the accelerating tube is directly below the terminal. The rings around the accelerating tube, connected by high resistances, help to ensure a uniform field, which reduces sparking.



78 Chapter 17 • Radioactivity and Nuclear Reactions

the accelerating fields, we refer to them as *AC accelerators*. (Since the frequency of the alternating voltage is usually in the range of 10 to 200 MHz, it is a radiofrequency and is often abbreviated RF.) Here we describe just two important types of AC accelerators, the linear accelerator and the cyclotron. We confine ourselves mostly to machines in which relativistic effects are not too large — proton energies up to a few hundred MeV, for example. A few of the more recent accelerators that produce particles with thousands and even millions of MeV are described in Section 18.11.

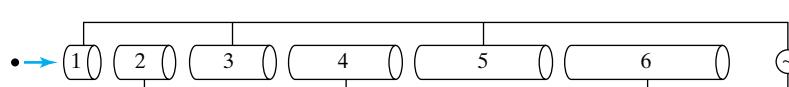
The Linear Accelerator

In the **linear accelerator**, or **linac**, each particle travels through a succession of many conducting cylinders, mounted in a single straight line (whence the name *linear accelerator*). Inside each cylinder, or *drift tube*, there is no electric field and the particle feels no force. However, the potentials of the tubes oscillate in such a way that whenever the particle reaches a gap between two tubes, there is an accelerating electric field in the gap. This field boosts the particle from one tube into the next, giving it an additional energy of V_0q , where V_0 is the potential difference across the gap. If the linac has N gaps in all, the particle gains an energy of NV_0q in its passage through the whole accelerator. In most linacs the particle is given a preliminary boost (by a Cockcroft–Walton accelerator, for example), so that it enters the first tube with kinetic energy K_0 and leaves the last with $K = K_0 + NV_0q$.

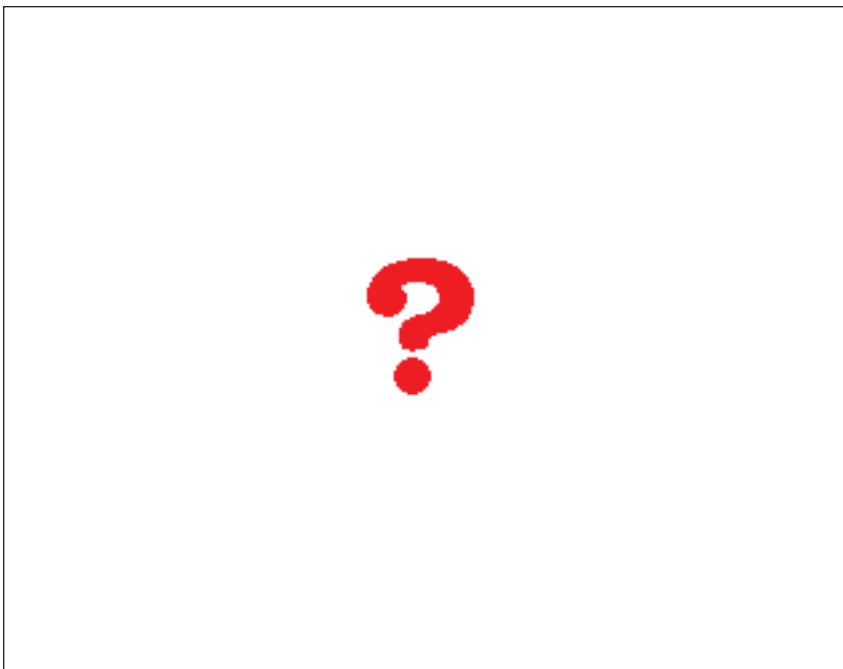
There are various ways to arrange for the required accelerating voltages in a linac. One simple method, used in some early machines, is sketched in Fig. 17.26, which shows a linac with just six cylindrical electrodes, and hence five accelerating gaps. An AC supply is connected to the six tubes, one side to tubes 1, 3, and 5, the other to tubes 2, 4, and 6. Thus the potential difference between successive tubes oscillates between two values $\pm V_0$, tubes 1, 3, 5 being at maximum potential when tubes 2, 4, 6 are at minimum and vice versa. Consider now a group of positive ions injected into tube 1 so that they arrive at the first gap just when tube 1 is at maximum potential (and, hence, tube 2 at minimum). These ions are accelerated from tube 1 to tube 2 by the potential difference V_0 , and each gains an energy V_0q .

Next, the length of tube 2 is chosen so that the time for the ions to traverse tube 2 is exactly one half the period of the AC supply.* Thus, by the time the ions reach the gap between tubes 2 and 3, tube 2 is at maximum potential, and tube 3 at minimum. Therefore, the ions receive a second boost of energy V_0q .

Each successive tube is designed in the same way, so that the ions take a half period to traverse the tube and hence arrive at the next gap at the moment of maximum potential drop. Since the ions are accelerating, this requires each tube to be somewhat longer than its predecessor. A typical frequency for the AC supply is of order 100 MHz, for which a half period is 5×10^{-9} s. In this



*Modern linacs work somewhat differently, using a traveling electromagnetic wave. (See Section 18.11.)

**FIGURE 17.27**

Section of the 800-MeV proton linac at Los Alamos, New Mexico. The picture shows the tank containing the drift tubes, which are not themselves visible. The entire linac is half a mile long.

time a proton with about 100 MeV, and hence $v \approx 0.5c$, travels nearly 1 m. Therefore, linear accelerators, sometimes with hundreds of tubes, can be very large machines. The 800-MeV proton linac at Los Alamos, shown in Fig. 17.27, is about half a mile long; the 50-GeV electron linac at Stanford is 2 miles long.

The Cyclotron

To reach useful energies, a linac must have many accelerating gaps. In 1929 it occurred to the American physicist Ernest Lawrence to use a magnetic field to make the accelerating particles move in circles. In this way the particles could be brought back repeatedly to the *same* accelerating gap.

Lawrence's accelerator, which came to be called the **cyclotron**, is shown schematically in Fig. 17.28. The magnetic field is produced by a magnet with two large circular pole faces. This can be seen in the side view of Fig. 17.28(a), with its uniform magnetic field pointing straight up between the poles. Also visible in the side view are the grounded vacuum chamber within which the ions are accelerated and an electrode, called a "dee," that accelerates the ions.*

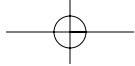
The dee is more clearly visible (and the origin of its name very apparent) in the overhead view of Fig. 17.28(b); it is a hollow semicircular conductor, inside which the ions spend half the period of their acceleration. The ions are produced in an electric discharge, close to the center of the dee; the discharge is fed by a gas whose nuclei are to be accelerated (hydrogen for protons, deuterium for deuterons, helium for alpha particles, and so on). The dee is connected to an AC voltage of amplitude V_0 , arranged so that the dee is at $-V_0$ and attracts the ions as they enter the dee, but is at $+V_0$ and repels them as they leave it. Thus the ions gain an energy V_0q each time they enter the dee

Ernest Lawrence (1901–1958, American)



Lawrence, who spent almost all of his career at Berkeley, invented the cyclotron, one of the most successful machines for accelerating subatomic particles to the energies needed to probe the atomic nucleus. This invention, which earned him the 1939 Nobel Prize in physics, was the precursor of the modern synchrotron, which can accelerate protons to energies of a million MeV.

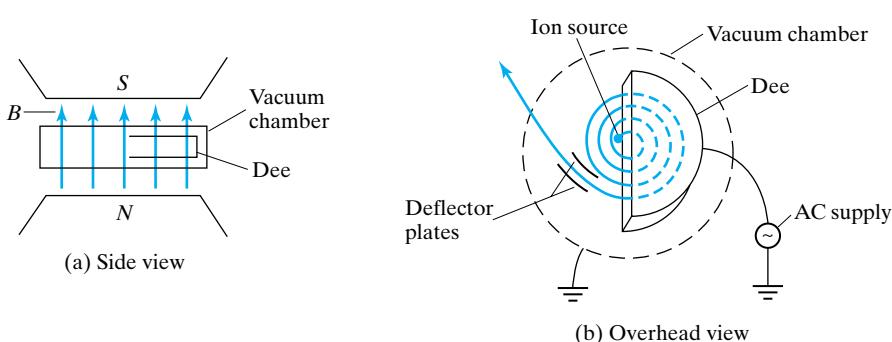
*Some cyclotrons have two or more electrodes to give a more rapid energy gain. For simplicity, we describe a single-electrode machine.



80 Chapter 17 • Radioactivity and Nuclear Reactions

FIGURE 17.28

The cyclotron. (a) Side view, showing the poles of the large magnet, the vacuum chamber in which the ions move, and the accelerating electrode called the “dee.” (b) Overhead view. The ions are produced at the center and gradually spiral outward. They gain an energy $V_0 q$ each time they enter the dee and again each time they leave it.



and again each time they leave it. Meanwhile the magnetic field forces the ions to follow a circular path with radius [Eq. (2.48)]

$$R = \frac{p}{qB} = \frac{mv}{qB} \quad (17.82)$$

(For the moment we assume the ions move nonrelativistically, so that $p = mv$.) Each time the ions speed up, this radius increases, so the ions spiral steadily outward from the ion source, as shown in Fig. 17.28. Eventually, when they approach the outer edge of the B field, a pair of charged plates deflects them outward to extract them from the machine. The photograph in Fig. 17.29 shows the dee of a cyclotron at the University of Colorado.

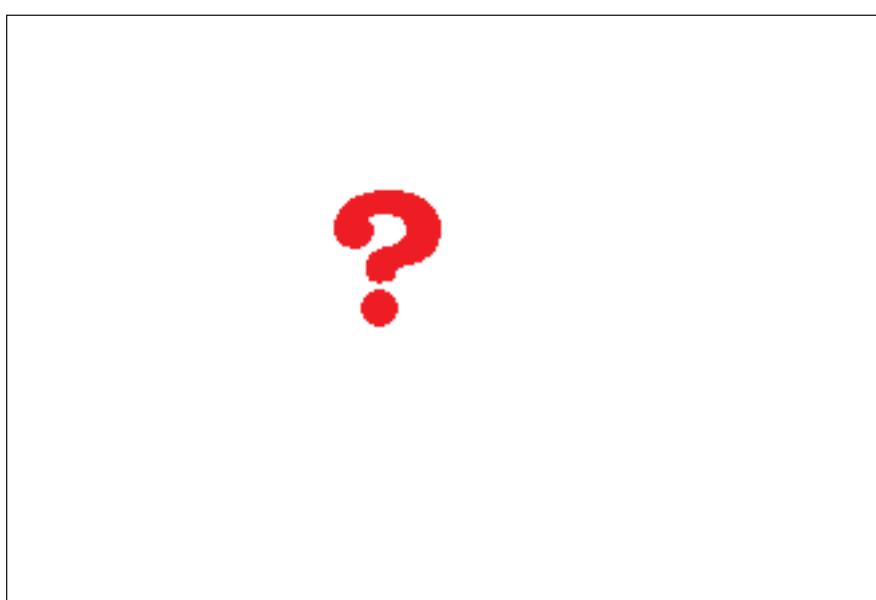
The frequency f with which the ions orbit is easily found from (17.82).

$$f = \frac{v}{2\pi R} = \frac{qB}{2\pi m} \quad (17.83)$$

The beautiful feature of this result — first appreciated by Lawrence — is that it is independent of v : With each boost, v increases but R also increases in direct proportion, so that the time for each orbit stays the same. Therefore, if we fix

FIGURE 17.29

The dee of the Colorado cyclotron was photographed while the machine was opened for maintenance. When in operation, the ion source was about 1 cm from the small metal block at the center of the dee, and the ions began their first orbit through the hole in this block.



the frequency of the AC supply at the value of the **cyclotron frequency** (17.83), the ions will continue to enter and leave the dee in time with the maximum accelerating field throughout the acceleration process. That a cyclotron can be driven by an AC supply of fixed frequency was a major simplification and made it a practical accelerator used in many nuclear laboratories around the world.

Example 17.6

A cyclotron accelerates protons to 10 MeV (and hence $v = 4.38 \times 10^7$ m/s) with a maximum orbital radius of 40 cm. **(a)** Find the strength B of the magnetic field. **(b)** What must be the frequency of the AC supply? **(c)** If the amplitude of the accelerating voltage is 50 kV, how many complete orbits are required to reach full energy?

- (a)** The maximum speed ($v = 4.38 \times 10^7$ m/s) is reached at the maximum radius ($R = 40$ cm). Thus, from (17.82),

$$B = \frac{mv}{qR} = \frac{(1.67 \times 10^{-27} \text{ kg}) \times (4.38 \times 10^7 \text{ m/s})}{(1.60 \times 10^{-19} \text{ C}) \times (0.40 \text{ m})} = 1.14 \text{ T}$$

- (b)** The frequency of the AC supply must equal the cyclotron frequency (17.83):

$$f = \frac{qB}{2\pi m} = \frac{(1.6 \times 10^{-19} \text{ C}) \times (1.14 \text{ T})}{2\pi \times (1.67 \times 10^{-27} \text{ kg})} = 17.4 \text{ MHz}$$

- (c)** A proton gains 50 keV each time it enters and each time it leaves the dee. Thus, in each complete orbit it gains 100 keV, and the total number of orbits is

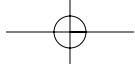
$$N = \frac{10 \text{ MeV}}{100 \text{ keV}} = 100 \text{ orbits}$$

In the simple form described here, the cyclotron depends crucially on the result (17.83) that the orbital frequency remains constant as the ions speed up. Unfortunately, this remains true only as long as the ions move nonrelativistically. Once they become relativistic, we must include a factor of γ in the expression (17.82) for R (since $p = \gamma mv$), and the frequency (17.83) becomes $f = qB/(2\pi m\gamma)$, which decreases like $1/\gamma$. As long as γ remains close to 1, this is a small effect that can be accommodated (for example, by arranging that B increases gradually with R), but for energies above several hundred MeV, new and more elaborate methods are needed, as we describe in Section 18.11.

17.12 Particle Detectors*

*This section could be omitted without loss of continuity.

Most nuclear measurements involve the detection of particles — particles ejected from radioactive nuclei, particles produced from accelerators to probe nuclei, and particles created in nuclear reactions. In addition to detecting these particles, one must usually measure some of their properties — their mass, charge, energy, momentum, and so on. In the business of detecting and



82 Chapter 17 • Radioactivity and Nuclear Reactions

measuring particles, there is a sharp distinction between charged and neutral particles. When a charged particle passes through matter (solid, liquid, or gas), it can ionize or excite many of the atoms that it passes; that is, it can eject electrons or raise them to excited states. This ionization or excitation is easily detected and is the basis for most detectors of charged particles. Neutral particles, such as the photon and neutron, are usually not as easily detected, and most detectors of neutral particles work by having the neutral produce a charged particle (as when a photon produces an electron in the photoelectric effect) and then detecting the charged particle.

Photographic Film

The discovery of radioactivity by Becquerel involved the darkening of photographic film by nuclear particles. Of course, the original purpose of photographic film was to detect visible photons and hence to record pictures. The essential component of photographic film is a silver halide* in the form of many small grains. When a photon passes through the film, it can knock an electron out of an atom in its path. This free electron initiates a sequence of events leading to the precipitation of silver atoms when the film is developed. This leaves the film darkened at those points where many photons passed through and produces a negative image of the original picture.

In much the same way, photographic film is sensitive to other kinds of photons, and it is, for example, widely used to detect X-ray photons for medical purposes. In the same way, also, film is sensitive to charged particles. Any charged particle passing through a film produces many free electrons, which cause the deposit of silver atoms and darken the film. This was the basis of Becquerel's discovery in 1896, and since that time, photographic film has been widely used to detect nuclear and subnuclear particles. In Chapter 16 (Fig. 16.18) we saw that it was used in early mass spectrometers to detect the arrival of ions. In Chapter 18 (Fig. 18.4) we show the paths of the nuclear particles called pions and muons recorded in photographic films. Photographic film can even be used to detect neutrons since the neutrons can knock protons out of water molecules in the film and the protons then darken the film in the usual way.

For all its versatility as a particle detector, photographic film has largely been replaced by various other kinds of detectors, several of which we describe here. One important application of photographic film that is still in wide use today is in the *film badge* worn by workers who may be exposed to nuclear radiation (Fig. 17.30). Film in these badges is removed periodically and developed to determine whether workers have been exposed to radiation. The darkness of the film is measured to monitor any exposure and to ensure that safety regulations are being followed properly.



FIGURE 17.30

A film badge used to monitor exposure to radiation.

Cloud and Bubble Chambers

Photographic film can produce a visible track of a nuclear particle, even though the particle itself is totally invisible. There are various other detectors that can do the same thing and are, in some ways, more convenient than photographic film. The first such was the **cloud chamber**, invented by the Scottish physicist C. T. R. Wilson in about 1910. This is a chamber containing air that is supersaturated with water vapor. When a charged particle passes through the chamber, water droplets condense around the ions produced by the particle.

*Silver bromide, silver chloride, or silver iodide.

These droplets form a visible trail, which can be photographed for later study. If the particle happens to collide with a nucleus in the chamber, any charged particles produced in the resulting nuclear reaction will also produce visible trails. Thus, many details of the reaction are revealed in a form that is easy to study. This was illustrated in Fig. 17.10, which is a photograph of the reaction $\alpha + {}^{14}\text{N} \rightarrow \text{p} + {}^{17}\text{O}$ in a cloud chamber.

A more recent variant of the cloud chamber is the **bubble chamber**, which contains superheated liquid hydrogen. The ions formed by any passing charged particle trigger boiling in the nearby liquid, producing tiny bubbles that give a visible record of the particle's path. One of the advantages of bubble chambers is that the density of nuclei in the liquid of the bubble chamber is some 1000 times that in the gas of a cloud chamber. Thus the probability that a passing particle will produce an interesting reaction in a bubble chamber is much higher than in a cloud chamber. Photographs of reactions in bubble chambers are reproduced in Figs. 18.5, 18.6, and 18.11.

If a magnetic field is applied to a cloud or bubble chamber, the path of any charged particle will be curved, and the particle's momentum can be found by measuring the radius of curvature [see Eq. (17.82)]. Further, the thickness of the trail left by a particle depends on its charge and mass, and under favorable conditions the particle can be completely identified by its track.

Ion Chambers

The **ion chamber** is a detector that makes direct use of the ionization produced by any charged particle. Figure 17.31 shows a simple ion chamber that contains two plates maintained at different potentials by a battery. Any charged particle passing through the air or other gas between the plates leaves a trail of electrons and positive ions. The electrons are drawn to the positive plate (or anode), and the positive ions are drawn to the negative plate (or cathode). For each electron that arrives at the anode, one negative charge flows around the circuit and neutralizes one positive charge at the cathode.

Although the current produced by the passage of a single charged particle is very small, it can be measured if amplified electronically. Thus the ion chamber can register the passage of individual charged particles, each of which produces a separate measurable pulse of current.

One convenient variant of the ion chamber is the **Geiger counter**, in which the two plates in Fig. 17.31 are replaced by a conducting cylinder and a wire running down its axis. The potential difference produces a large electric field near the narrow wire. When a charged particle ionizes an atom, some of the resulting charges are drawn near to the wire, where they accelerate rapidly and ionize more atoms. This chain reaction quickly produces a current that is large enough to detect without expensive amplifiers and can, for example, be made to cause an audible click in a small loudspeaker. Thus the Geiger counter is an economical, and often portable, detector of nuclear radiation.

In addition to detecting particles, an ion chamber can measure their energies. When a charged particle ionizes an atom, it loses an energy that has, on average, the same value whatever the particle's speed. Therefore, the total energy that the particle loses in the chamber is proportional to the number of atoms ionized, and hence to the total charge that flows in the circuit. In particular, if the chamber is large enough, the particle will come to rest, and, by measuring the charge flow, one can find the total kinetic energy with which the particle entered the chamber.

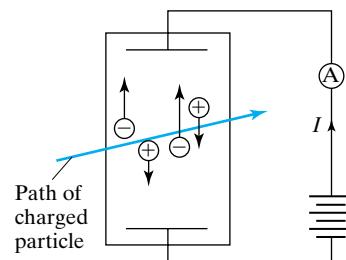
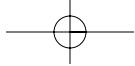


FIGURE 17.31

Schematic sketch of an ion chamber. A charged particle passing through the gas in the chamber produces a trail of electrons and positive ions, which are drawn to the plates and cause a measurable current I in the circuit.



84 Chapter 17 • Radioactivity and Nuclear Reactions

Example 17.7

Alpha particles ejected in the radioactive decay of ^{235}U are detected in an ion chamber. Each atom ionized costs an alpha particle about 30 eV of energy. Given that the alphas were ejected with about 4.4 MeV and assuming that they come to rest in the chamber, find the total charge that flows in the circuit when one particle is detected.

The total number of atoms ionized is just the incident energy of the alpha particle divided by the energy lost in each ionization.

$$N = \frac{4.4 \times 10^6 \text{ eV}}{30 \text{ eV}} = 1.5 \times 10^5$$

The total charge flowing around the circuit is therefore

$$Q = Ne = 2.4 \times 10^{-14} \text{ C}$$

With modern electronic amplifiers one can measure such charges with an accuracy of about 10^{-17} C . Thus one can measure Q , and hence the particle's original energy, with an accuracy of about 1 in 1000; that is, the kinetic energy can be measured within a few keV.

To measure a particle's energy with an ion chamber, one must be sure that the particle stops before it can leave the chamber. The higher a particle's energy, the farther it can travel before coming to rest; and for particles with many MeV, this requires the gas-filled chamber to be inconveniently large. One way around this difficulty is to replace the gas with a solid: Because the densities of solids are some 1000 times greater than those of gases, the particle meets 1000 times more atoms (in a given distance) and therefore loses all its energy in one thousandth of the distance. We next describe two detectors that use a solid medium: solid-state and scintillation detectors.

Solid-State Detectors

One cannot simply replace the gas of an ion chamber by any solid. If the solid is an insulator, the charges produced by ionization cannot flow to the collecting plates; if the solid is a conductor, a current will flow all the time, making it difficult to detect the small extra current caused by a passing particle. Fortunately, certain semiconductors can be arranged to act as insulators *except* when a charged particle passes through them. (Semiconductors were discussed in Section 14.2.) By placing a suitable semiconductor between two collecting plates, one can make a **solid-state detector** that acts much like a gas-filled ion chamber, but can stop a high-energy particle — and hence measure its energy — in a much smaller volume.

Scintillation Detectors

When charged particles pass through matter, they not only ionize atoms; they also elevate atoms to excited states. These excited atoms then give off light as they fall back to the ground state, and this light is exploited in the **scintillation detector**. One of the earliest such detectors was the zinc sulfide screen used by Rutherford in many of his experiments with alpha particles. Each time an alpha hit the screen, the tiny flash of light that it produced was observed and

recorded by the experimenter — a tedious and tiring job, which could be done only in a totally dark room.

Today, the light from a scintillation detector is monitored automatically. A photoelectric cell converts the light into an electric pulse, which, amplified by a photomultiplier if necessary, can be fed directly into a computer for processing and recording.

Most modern scintillation detectors use materials such as NaI crystals and certain plastics that are transparent to the light which they produce. This lets one use a block of material that is thick enough to stop the particles and hence to measure their energy. Some detectors use a liquid scintillator, which allows a radioactive sample to be placed right inside the scintillator, to improve the chance of detecting weak or low-energy signals.

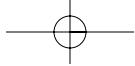
Wire Chambers

Ion chambers, solid-state detectors, and scintillation detectors all produce electric signals that can be fed directly to a computer. In addition to detecting a particle, they can also measure its energy, *provided* that the particle comes to rest in the detector. However, many modern experiments use particles with thousands and even millions of MeV, and such particles cannot be stopped in any detector of reasonable size. Instead, their energy is usually found by measuring the curvature of their path in a magnetic field, and this requires a detector that records the path of the particle. We have already described three such detectors — photographic films, cloud chambers, and bubble chambers — but all of them record their information in the form of a picture, which has to be interpreted and measured by a human experimenter. The **wire chamber** is a detector that came into use in the 1970s and records a particle's trajectory directly in the form of electrical pulses.

In a wire chamber many closely spaced, parallel wires are positively charged to attract the electrons produced by a fast ionizing particle. The wire closest to the path of the particle captures most of the electrons and produces a signal to record this point on the path. A second grid with its wires perpendicular to the first allows both x and y coordinates of the path to be measured. Several such pairs of wire grids determine the particle's entire trajectory through a magnetic field, and from this its momentum can be calculated.

Neutral Particles

When a charged particle such as a proton passes through matter, its electric field pulls on the electrons in the atoms along its path and leaves a trail of electrons, ions, and excited atoms. This is the basis for all the detectors described in this section. Neutral particles do not produce trails of ionization and excitation, but they can occasionally collide with atoms and produce moving charged particles. These charged particles can then be detected in the usual way. To detect neutrons, for example, hydrogen (in one form or another) is introduced into a detector; some of the neutrons collide with the hydrogen nuclei (protons) and the recoiling protons are detected as usual. Photons are detected by three main mechanisms, all of which produce moving electrons: (1) the Compton effect, in which the photon scatters off an atomic electron, causing it to recoil; (2) the photoelectric effect, in which a photon is absorbed and gives its energy to an electron; and (3) pair production, in which a photon's energy creates an electron-positron pair. A cloud-chamber photograph of the latter process is reproduced in Fig. 18.1, which shows clearly the absence of any visible track for the photon, but two clear tracks for the electron and positron.



86 Chapter 17 • Radioactivity and Nuclear Reactions

17.13 Units of Radiation*

*This section could be omitted without loss of continuity.

There are various units used to measure amounts of radiation. These units are especially important in radiology, the study of the effects of radiation on biological systems. Radiation has been used successfully to treat several diseases, particularly cancer, and radioactive “tracers” have proved invaluable in following the paths of biological processes. However, excessive exposure to radiation can cause serious illness and even death. Therefore, safety regulations are needed for radiologists and other people exposed to radiation. To formulate these regulations, one must have units with which to measure the radiation and its effects.

The various units of radiation fall into two main classes: (1) units of *source activity*, which characterize the radiation *produced* by a radioactive source, and (2) units of *absorbed radiation*, which measure the *effects* of radiation on substances exposed to it. As in most areas of physics, the units defined by early workers in the field have been replaced by internationally accepted SI units. However, because the older units are still widely used, we describe both the SI units and their older counterparts.

Units of Source Activity

The SI unit that measures the amount of radioactivity in a source is the **becquerel**, or Bq:

1 Bq of a radioactive substance is that amount
which produces 1 decay per second

As an example of the use of this unit, consider the naturally occurring radioactive gas radon 222, whose decay constant is $r = 2.1 \times 10^{-6} \text{ s}^{-1}$. To get one decay per second, we would need

$$N = \frac{1}{2.1 \times 10^{-6}} = 4.8 \times 10^5$$

atoms of radon. Therefore, 1 Bq of radon consists of 4.8×10^5 atoms and has a mass

$$(\text{mass of 1 Bq of radon}) = (4.8 \times 10^5) \times (222 \text{ u}) = 1.8 \times 10^{-19} \text{ kg}$$

Recent studies of the radon that is released by natural radioactivity in the ground and seeps into the basements of houses suggest that in the average U.S. home, the level of radon is about 50 Bq/m^3 . This is disturbingly close to the level of 150 Bq/m^3 above which the U.S. Environmental Protection Agency recommends remedial action because of increased risk of cancer.

The earlier unit of activity, still in wide use today, was the *curie*, or Ci, which is now defined in terms of the Bq as

$$1 \text{ Ci} = 3.7 \times 10^{10} \text{ Bq} = 3.7 \times 10^{10} \text{ decays/second}$$

The origin of this definition is that 1 Ci is, at least approximately, the activity of 1 gram of radium 226.

Units of Absorbed Radiation

The becquerel measures the activity of a radioactive source, but the biological effect of radiation depends more on the amount, or *dose*, that is absorbed by a living tissue. The **absorbed dose** resulting from any exposure to radiation is defined as the energy absorbed per unit mass of the exposed tissue. The SI unit for absorbed dose is the gray,

$$1 \text{ Gy} = 1 \text{ gray} = 1 \text{ J/kg}$$

that is, if the tissue absorbs an energy of 1 joule per kilogram of tissue, we say that it has absorbed a dose of 1 gray. The older unit for absorbed dose is the rad* (radiation absorbed dose), defined as

$$1 \text{ rad} = 0.01 \text{ Gy} = 0.01 \text{ J/kg}$$

The damage caused to living tissue by a given dose of radiation depends on the nature and energy of the radiation. For example, exposure to neutrons with energies of a few hundred keV can cause about 10 times the damage produced by the same dose of X-rays. The **relative biological effectiveness**, or **RBE**, of any radiation is defined by comparing its effects with those of 200-keV X-rays. Specifically, the RBE of a given type of radiation is the factor by which the damage it causes exceeds that of 200-keV X-rays. For example, the RBE of alpha particles is about 20, which means that a given dose of alpha radiation causes 20 times as much damage as the same dose of 200-keV X-rays. Table 17.2 shows approximate values for the RBE (also called the quality factor, QF) for some types of radiation.

TABLE 17.2

The relative biological effectiveness, or RBE, of some types of radiation.

Type of Radiation	RBE
Photons (200-keV X-ray)	exactly 1
Photons (4-MeV γ -ray)	0.6
Electrons	about 1
Neutrons	2 to 10
Protons	about 10
Alpha particles	up to 20

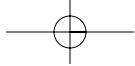
The biological damage caused by an exposure to radiation is determined by the product of the absorbed dose and the RBE of the radiation concerned. This product is called the **effective dose**.

$$\text{effective dose} = (\text{absorbed dose}) \times \text{RBE}$$

The SI unit for effective dose is called the sievert or Sv; thus

$$(\text{effective dose in Sv}) = (\text{absorbed dose in Gy}) \times \text{RBE}$$

*An earlier unit was the *roentgen*, which was about 0.87 rad. This was the dose that produced 1 electrostatic unit of ions ($3.3 \times 10^{-10} \text{ C}$) in 1 cm^3 of air.



88 Chapter 17 • Radioactivity and Nuclear Reactions

With this definition, 1 sievert of one kind of radiation causes roughly the same damage as 1 sievert of any other kind. The corresponding older unit is the **rem** (rad equivalent in man), so that

$$(\text{effective dose in rem}) = (\text{absorbed dose in rad}) \times \text{RBE}$$

Most normal exposures to radiation result in effective doses that are conveniently measured in millisieverts (mSv). The effective dose resulting from a single medical X-ray is of order 0.1 mSv. The annual effective dose from cosmic rays and other natural sources of radiation is of order 1 mSv for the average person living at sea level. (In the Rocky Mountain states, this number is nearer to 2 mSv because of the higher intensity of cosmic rays and the greater concentration of natural radioactivity in the ground.) These figures for natural exposure do not include the effects of breathing natural radon gas into the lungs; it has recently been realized that this may well contribute an annual dose of another 2 or 3 mSv.

Catastrophic exposures are typically thousands of mSv. An effective dose of 5000 mSv to the whole body at one time is fatal in about 50% of cases. A single exposure above 1000 mSv is known to increase a person's chance of fatal cancer by about 1.5×10^{-5} per mSv of exposure. In setting safety standards, it is generally assumed that this same rate applies at low levels of radiation, although there is substantial evidence that at lower doses the rate is actually less.*

The main units of radiation are summarized in Table 17.3.

TABLE 17.3

Principal units of radiation.

Quantity	SI unit	Traditional unit
Source activity	Bq (becquerel) = 1 decay/second	Ci (curie) = 3.7×10^{10} Bq
Absorbed dose	Gy (gray) = 1 J/kg	Rad (radiation absorbed dose) = 10^{-2} Gy
Relative biological effectiveness	RBE	RBE
Effective dose (eff dose) = (abs dose) × RBE	Sv (sievert) (eff dose in Sv) = (abs dose in Gy) × RBE	Rem (rad equivalent in man) = 10^{-2} Sv (eff dose in rem) = (abs dose on rad) × RBE

CHECKLIST FOR CHAPTER 17

CONCEPT	DETAILS
Radioactivity	Sec. 17.2
α	$(Z, N) \rightarrow (Z - 2, N - 2) + \alpha$ (17.1)
β	$(Z, N) \rightarrow (Z + 1, N - 1) + e + \bar{\nu}$ (17.23)
γ	excited nucleus \rightarrow nucleus + photon

*Report of the Committee on the Biological Effects of Ionizing Radiation, National Academy of Sciences, Washington, DC (1988).

Exponential decay	$N(t) = N_0 e^{-rt}$ (17.6)
decay constant	r = probability per time of decay of one nucleus
mean life	$\tau = 1/r$ (17.7)
half-life	$t_{1/2} = \tau \ln 2$ (17.11)
Radioactive dating	Sec. 17.3
β^\pm decay and electron capture	Sec. 17.4
Neutrinos	Neutral, very low-mass particle emitted in β decay
Natural radioactive series	Three natural sequences of radioactive chains (Sec. 17.5)
Valley of stability	Curve in (Z, N) of most stable nuclei (Fig. 17.7)
Nuclear reactions	Nuclear transformations when two nuclei collide
conservation of nucleon number	$\sum A_{\text{initial}} = \sum A_{\text{final}}$ (17.33) AU: OK
conservation of charge	$\sum q_{\text{initial}} = \sum q_{\text{final}}$ (17.34)
cross section σ	Effective area of target for projectile (Sec. 17.6)
Nuclear fission	Energy gain when a heavy nucleus splits (Sec. 17.7)
chain reaction	Neutrons from one fission go on to cause more fissions.
Nuclear fusion	Energy gain when two light nuclei fuse. (Sec. 17.8)
fusion in stars	Proton-proton cycle
Theory of α decay*	Quantum tunneling lets alpha particles escape (Sec. 17.10)
Accelerators*	DC, linear, cyclotron (Sec. 17.11)
Detectors*	Film, cloud and bubble chamber, ion chamber, scintillators, wire chambers (Sec. 17.12)
Units of radiation*	Activity and dose (Sec. 17.13)

PROBLEMS FOR CHAPTER 17

SECTION 17.3 (The Exponential Decay Law)

- 17.1** • The decay constant of ^{210}Rn is $r = 0.29 \text{ hour}^{-1}$. If a sample of ^{210}Rn contains 10^{12} atoms at $t = 0$, how many will remain: (a) after 1 hour, (b) after 1 day, (c) after 2 days?
- 17.2** • The decay constant for a certain nucleus is $r = 0.010 \text{ min}^{-1}$. At $t = 0$ a sample contains 10^6 of these nuclei. (a) How many of these nuclei will remain after an hour? (b) What is the total rate of decays, $R(t)$, at $t = 0$? (c) What is $R(t)$, after an hour?
- 17.3** • (a) Use the data in Appendix D to find the decay constant and mean life of uranium 232. (b) Do the same for polonium 214.
- 17.4** • The isotope cobalt 60 has a decay constant $r = 4.2 \times 10^{-1} \text{ s}^{-1}$. Consider a pure sample of $1 \mu\text{g}$ of ^{60}Co . (a) How many atoms are in this sample? (b) How many ^{60}Co atoms will remain after one year? (c) What is the half-life of ^{60}Co ?
- 17.5** • A sample of a certain kind of nucleus has mean life equal to 10 minutes. (a) What is its decay constant r ? (b) Use Equation (17.6) to find the time t at which the
- number of these nuclei will have dropped to half its initial value.
- 17.6** • At time $t = 0$ a sample contains 600 atoms of phosphorus 32. How many atoms of phosphorus 32 would remain a month (30 days) later?
- 17.7** • (a) Given that ^{238}U has a half-life of 4.47×10^9 years, find its decay constant r . (b) How many atoms are there in 1 g of pure ^{238}U ? (c) How many decays are there per second in 1 g of ^{238}U ?
- 17.8** • The nucleus of polonium 208 has a decay constant $r = 0.239 \text{ year}^{-1}$. (a) What is the half-life of ^{208}Po ? (b) How many half-lives are there in 9 years? (c) If a sample of ^{208}Po contains 8000 atoms today, how many will remain after 9 years?
- 17.9** • From the exponential decay law (17.6), prove that
- $$\frac{N(t + \Delta t)}{N(t)} = e^{-r\Delta t}$$

that is, the fractional decrease in N during any time interval from t to $t + \Delta t$ depends on Δt , but not on t itself.

90 Chapter 17 • Radioactivity and Nuclear Reactions

17.10 •• The decay constant of a certain nucleus is 0.05 s^{-1} . At $t = 0$ a sample contains 6×10^{15} nuclei. **(a)** How many nuclei remain at $t = 10 \text{ s}$? **(b)** How many remain at $t = 60 \text{ s}$ and at 70 s ? **(c)** Evaluate the ratios N_{10}/N_0 and N_{70}/N_{60} , and explain why they are equal.

17.11 •• Prove that the exponential decay law $N(t) = N_0 e^{-rt}$, can be rewritten as $N(t) = N_0 2^{-t/t_{1/2}}$, where $t_{1/2} = \ln(2)/r$. [Hint: $e^{\ln(2)} = 2$.]

17.12 •• In Section 17.3 we proved that the number $N(t)$ of radioactive nuclei surviving after a time t satisfies the equation $dN/dt = -rN$, where r is the decay constant of the nuclear species in question. Rewrite this equation as $dN/N = -r dt$, and by integrating, prove that $N = N_0 \exp(-rt)$, where N_0 is any constant.

17.13 •• Half of the nuclei in a sample decay before the time $t = t_{1/2}$, and the other half after; that is, $t_{1/2}$ is the *median* life. On the other hand, $t_{1/2}$ is *not the mean* (or average) life. To find the mean life, note that the number of nuclei that decay between times t and $t + dt$ is $R(t) dt$. Thus, the fraction of the original N_0 nuclei that live for a time between t and $t + dt$ is $R(t) dt/N_0$. If you multiply this fraction by t and integrate over all possible values of t , you will obtain (by definition of the average) the average life $\langle t \rangle$.

$$\langle t \rangle = \int_0^\infty \frac{t R(t)}{N_0} dt$$

Evaluate $\langle t \rangle$ in terms of the decay constant r . How is $\langle t \rangle$ related to the lifetime τ defined in Eq. (17.7)?

17.14 •• At $t = 0$ a sample contains $N_0 = 64,000$ radioactive nuclei, all of the same species. If the decay constant is $r = 0.01 \text{ s}^{-1}$, how many of the nuclei remain after **(a)** 2 min, and **(b)** 20 min? **(c)** Is your answer in part (b) the actual number of nuclei you would expect to find after 20 min? Explain. (Remember that a nucleus is a quantum system, and quantum mechanics deals in probabilities.)

17.15 •• A radioactive nucleus A has decay constant r and decays to a stable offspring nucleus B . At $t = 0$, a pure sample of A contains N_0 atoms. Derive an expression for the number $N_B(t)$ of offspring nuclei in the sample at time $t \geq 0$. Sketch $N_B(t)$ as a function of t .

17.16 •• A radioactive nucleus A has decay constant r_A . It decays to an offspring nucleus B , which decays much more rapidly, with decay constant $r_B \gg r_A$. The population of A decreases exponentially as usual [$N_A(t) = N_0 \exp(-r_A t)$]. Because B decays so rapidly, each B decays almost as soon as it is produced. Therefore, the rate of production of the offspring B (by decay of the parents A) is approximately equal to the rate of loss (by decay of the B nuclei themselves). Use this condition to find an expression for $N_B(t)$ in terms of N_0 , r_A , r_B , and t . Show in particular that N_B remains proportional to N_A .

17.17 •• The carbon in living organisms contains about 1.3 atoms of ^{14}C for every 10^{12} atoms of all kinds of car-

AU: are
the sub-
scripts
"B"
roman?

bon, and the half-life of ^{14}C is 5730 years. **(a)** How many decays would one expect per minute from 7 g of carbon taken from a living tree? **(b)** If 7 g of carbon is extracted from a wooden beam of a prehistoric dwelling and if they undergo 10 decays per minute, about how old is the dwelling?

17.18 •• **(a)** At what rate would 2 g of carbon from a living organism eject electrons? **(b)** If archeologists extract 2 g of carbon from a dog's skull and count 21 electrons ejected per minute, how long ago did the dog die?

SECTION 17.4 (Beta Decay and the Neutrino)

17.19 • Equation (17.22), which gives the kinetic energy released in β^- decay, was derived on the assumption that there are only two final particles (the electron and the residual nucleus, but no neutrino). Prove, nevertheless, that (17.22) correctly predicts the total kinetic energy available to all three final particles, given that the neutrino's mass is negligible in this context.

17.20 • Use the data in Appendix D to find the total kinetic energy released in the β^- decay of ^{14}C . (See Problem 17.19.)

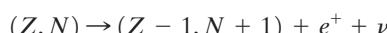
17.21 • Calculate the total kinetic energy released in the β^- decay of a single free neutron.

17.22 •• Write down an equation [similar to Eq. (17.25)] to describe the β^- decay of each of the following nuclei: **(a)** ^{60}Co , **(b)** ^3H (tritium), **(c)** ^{15}O , and **(d)** the decay by electron capture of ^{231}U .

17.23 •• If we neglect the neutrino's mass, the total kinetic energy, K , released in the β^- decay $(Z, N) \rightarrow (Z + 1, N - 1) + e^- + \bar{\nu}$ is given by (17.22). In particular, the criterion that β^- decay is possible is just that K must be positive, or

$$m_{\text{atom}}(Z, N) \geq m_{\text{atom}}(Z + 1, N - 1) \quad (\text{for } \beta^- \text{ decay})$$

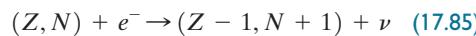
In other words, one has only to compare the atomic masses involved to see if β^- decay is possible. **(a)** Find an expression for the total kinetic energy released in the β^+ decay



and show that the condition for this process to be possible is

$$m_{\text{atom}}(Z, N) \geq m_{\text{atom}}(Z - 1, N + 1) + 2m_e \quad (\text{for } \beta^+ \text{ decay}) \quad (17.84)$$

(b) Do the same for the electron capture



assuming that the original two particles are at rest, and show that the condition for electron capture is the same as (17.84) *except* that the term $2m_e$ is missing. **(c)** The mass of ^{231}U is 231.03626 u, while that of ^{231}Pa is 231.03588 u. Show that ^{231}U can decay by electron capture but not by β^+ emission.

17.24 ••• Observations show that if there were no neutrino, β decay would violate conservation of momentum, as well as energy. For example, in a decay of ${}^6\text{He}$ at rest, the two final charged particles were observed with momenta as shown in Fig. 17.32. **(a)** Show that this decay would violate conservation of momentum if there were no neutrino involved. **(b)** Assuming that a neutrino is ejected and that total momentum is conserved, calculate the momentum (magnitude and direction) of the unobserved neutrino. **(c)** Assuming conservation of energy, find the neutrino's energy. **(d)** Show that your answers in parts (b) and (c) are consistent with the neutrino's having zero rest mass (within the accuracy of the given numbers).

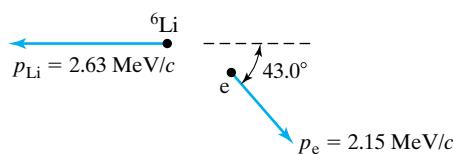


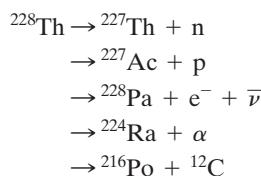
FIGURE 17.32
(Problem 17.24)

17.25 ••• **(a)** Use the semiempirical binding-energy formula (16.30) to write an expression for the mass of a nucleus in terms of Z and A . **(b)** Now consider a set of isobars with A odd. Use your mass formula to show that a plot of mass against Z , as in Fig. 17.2, should fit a parabola ($m = \alpha Z^2 + \beta Z + \gamma$). **(c)** Now consider a set of isobars with A even. Show that because of the pairing energy, the masses will alternate between two parabolas, one for the nuclei with Z even and the other for Z odd. **(d)** Sketch the plot of m against Z , and explain why one can have two stable isobars if A is even. **(e)** Find two examples of this phenomenon in Appendix D.

17.26 ••• In an observed β decay $X \rightarrow Y + e^- + \bar{\nu}$ of a nucleus X that is exactly at rest, it happens that the electron is also exactly at rest. For this special case, derive an exact (relativistic) expression for the energy of the $\bar{\nu}$ in terms of m_X , m_Y , and m_e (assuming that the mass of the neutrino is negligible). Show that to an excellent approximation, the neutrino takes all of the available kinetic energy, ΔMc^2 , when the electron is at rest. In a typical case [$m_X \approx 200$ u and $(m_X - m_Y) \approx 0.001$ u] how good is this approximation? (Give your answer as a fractional error.)

SECTION 17.5 (The Natural Radioactive Series)

17.27 • Use the data in Appendix D to find the kinetic energy released in each of the following conceivable decays:



Which of these processes is possible? (Note: You will find that the last one is possible. However, its probability is so small compared to α decay that it has not been observed.)

17.28 •• The supernova explosion that is believed to have preceded the formation of our solar system created large quantities of radioactive nuclei. There are reasons to think that the isotopes ${}^{235}\text{U}$ and ${}^{238}\text{U}$ would have been created in roughly equal numbers. That is, the ratio N_{238}/N_{235} was close to 1 at $t = 0$ (the time of the explosion). **(a)** Show that at any later time

$$\frac{N_{238}}{N_{235}} = \exp[(r_{235} - r_{238})t]$$

where r_{235} and r_{238} denote the decay constants of ${}^{235}\text{U}$ and ${}^{238}\text{U}$. **(b)** Use the data in Appendix D to find r_{235} and r_{238} . **(c)** Use the observed abundances of ${}^{235}\text{U}$ and ${}^{238}\text{U}$ on earth today to estimate how long ago the supernova explosion occurred.

17.29 •• **(a)** Use the data in Appendix D to calculate the total kinetic energy released in the α decay of ${}^{220}\text{Rn}$. **(b)** Find the total kinetic energy released in the β decay of ${}^{220}\text{Rn}$. Is this decay possible? **(c)** Answer the same questions for ${}^{212}\text{Bi}$. **(d)** What is the relevance of your answers in consideration of Fig. 17.9?

17.30 •• Use the data in Appendix D to write down the whole of the radioactive series that begins with ${}^{238}\text{U}$. Draw the series on a plot of Z against N as in Fig. 17.9.

17.31 •• Use the data in Appendix D to write down the whole of the radioactive series that begins with ${}^{235}\text{U}$. Draw the series on a plot of Z against N as in Fig. 17.9.

17.32 •• The mass number A decreases by 4 in α decay, but is unchanged in β or γ decay. This means that throughout a radioactive series A changes only in increments of 0 or 4. Thus, if A happens to be a multiple of 4 ($A = 4n$) in the original parent, it will remain a multiple of 4 in all members of the series. This is what happens in the ${}^{232}\text{Th}$ series, which is therefore called the $4n$ series. In the same way, there are three other possible series: $(4n+1)$, $(4n+2)$, and $(4n+3)$. **(a)** What are the original parents of the $(4n+2)$ and $(4n+3)$ series? **(b)** Use the data in Appendix D to explain why the $(4n+1)$ series is not found naturally.

17.33 •• **(a)** Use the data in Appendix D to predict the total kinetic energy released in the α decay of ${}^{227}\text{Th}$. **(b)** Figure 17.33 shows the observed energy spectrum in the α decay of ${}^{227}\text{Th}$ nuclei (all in their ground

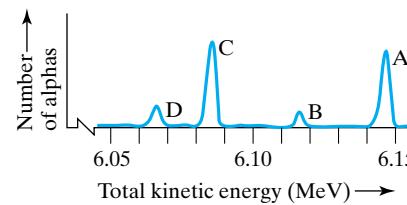


FIGURE 17.33
(Problem 17.33)

92 Chapter 17 • Radioactivity and Nuclear Reactions

state). Can you explain the three spikes labeled B, C, and D? (c) Use Fig. 17.33 to predict the excitation energy of three excited states of ^{223}Ra .

- 17.34 ••** Consider the α decay of a nucleus X at rest, $X \rightarrow Y + \alpha$, where Y denotes the offspring nucleus. (a) Prove that because the α is much lighter than Y, almost all the available kinetic energy goes to the α ; specifically, prove that

$$K_\alpha = \frac{m_Y}{m_\alpha} K_Y \quad (17.86)$$

[Hint: Use conservation of momentum and treat all particles nonrelativistically.] (b) In the α decay of a typical naturally radioactive nucleus, what percentage of the available kinetic energy does the alpha particle take? (c) Assuming that β decay has the form $X \rightarrow Y + e$, use the relation corresponding to (17.86) to find what percentage of the available kinetic energy goes to the electron. (Your answer here is really an upper bound, since some energy can go to the neutrino. It is also only approximate since the electron can be fairly relativistic.)

SECTION 17.6 (Nuclear Reactions)

- 17.35 •** In most nuclear reactions the total nucleon number is conserved; that is, $\sum A_{\text{initial}} = \sum A_{\text{final}}$, where $A = Z + N$ for each nucleus, as usual. Show that, except in reactions involving β decay, both $\sum Z$ and $\sum N$ are separately conserved. [Hint: Total charge is always conserved, and except in β decay, all the charge is carried by protons.]

- 17.36 •** Use Eq. (17.37) and the data of Appendix D to predict ΔK , the kinetic energy released in the reaction $p + ^7\text{Li} \rightarrow \alpha + \alpha$.

- 17.37 •** When Rutherford first observed the reaction (17.30), he assumed that the alpha particle simply knocked a proton out of the ^{14}N in the process $\alpha + ^{14}\text{N} \rightarrow \alpha + p + ^{13}\text{C}$. Given that the incident alpha particles had about 5.6 MeV, was this reaction possible?

- 17.38 •** Before two nuclei can react, they must come within 1 fm or so of one another, and because of their Coulomb repulsion, this requires a lot of energy. (a) Calculate the energy needed to bring a proton from infinity to a separation of 1 fm from an alpha particle. (1 fm = distance from surface to surface.) (b) Do the same for an alpha particle and a uranium nucleus. [Use Eq. (16.3) to estimate the radii involved.]

- 17.39 •** To find the mass of the neutron, Chadwick used the reaction



The best available values for the masses of the three nuclei were $m_\alpha = 4.00106 \text{ u}$, $m_B = 11.00825 \text{ u}$, and $m_N = 14.0042 \text{ u}$. The kinetic energies Chadwick measured as $K_\alpha = 5.26 \text{ MeV}$, $K_B = 0$, $K_n = 3.26 \text{ MeV}$, and $K_N = 0.57 \text{ MeV}$. Using these data, find Chadwick's value for the mass of the neutron, and compare with the current value.

- 17.40 •** The cross section for 0.01-MeV neutrons to be captured by uranium 238 is 0.7 barn. Two grams of ^{238}U is exposed to a flux of 3×10^9 neutrons per square meter (with energy 0.01 MeV). How many atoms of ^{239}U will be produced?

- 17.41 •** A hunter observes 100 crows settling randomly among the leaves of a large oak tree, where they are no longer visible. He fires 500 shots at random into the tree and hits 25 of the crows. If the area of the tree (as seen by the hunter) is 1200 ft^2 , what is σ , the cross-sectional area of a single crow? (Notice how, in a classical experiment of this type, σ is the actual cross-sectional area of the target.)

- 17.42 •** Which of the following reactions violate the laws of conservation of charge or nucleon number (and are therefore impossible)?

- (a) $p + ^{63}\text{Cu} \rightarrow n + ^{63}\text{Ni}$
- (b) $n + ^{64}\text{Zn} \rightarrow ^2\text{H} + ^{63}\text{Cu}$
- (c) $n + ^{56}\text{Fe} \rightarrow \alpha + ^{53}\text{Mn}$
- (d) $n + ^{14}\text{N} \rightarrow p + ^{13}\text{C}$

- 17.43 ••** Use the laws of conservation of energy, charge, and nucleon number to decide which of the following processes are impossible. Give your reasons.

- (a) $^{207}\text{Po} \rightarrow ^{207}\text{At} + e^- + \bar{\nu}$
- (b) $^{213}\text{Rn} \rightarrow ^{208}\text{Po} + \alpha$
- (c) $^{212}\text{Po} \rightarrow ^{208}\text{Pb} + \alpha$
- (d) $p + ^{14}\text{N} \rightarrow \alpha + ^{12}\text{C}$
- (e) $n + ^{22}\text{Na} \rightarrow p + ^{22}\text{Ne}$

- 17.44 ••** Reactions in which kinetic energy is lost ($\Delta K < 0$) are called **endothermic** and cannot occur below a minimum (threshold) initial energy. (Reactions in which kinetic energy is released are called **exothermic**.) Which of the following reactions are endothermic, and what are their threshold kinetic energies?

- (a) $n + ^{208}\text{Pb} \rightarrow p + ^{208}\text{Tl}$
- (b) $n + ^{207}\text{Pb} \rightarrow \alpha + ^{204}\text{Hg}$
- (c) $n + ^{187}\text{Os} \rightarrow p + ^{187}\text{Re}$
- (d) $n + ^{197}\text{Au} \rightarrow ^2\text{H} + ^{196}\text{Pt}$

- 17.45 ••** When the Joliot-Curies observed the reaction $\alpha + ^9\text{Be} \rightarrow n + ^{12}\text{C}$, they suggested (erroneously) that the neutral particle was a photon. Chadwick's main argument against this concerned the large momentum (up to $100 \text{ MeV}/c$) of the protons that these supposed photons ejected from wax. Assuming that it was photons that ejected the protons and given that a proton gets the largest momentum in a head-on collision, do the following: (a) To a fair approximation, the photon would bounce back with its energy unchanged; in this approximation, what is the incident photon energy required to eject a proton with momentum $100 \text{ MeV}/c$? (b) Using conservation of energy and momentum, redo this calculation without the approximation of part (a).

17.46 ••• One of the ways that Chadwick estimated the neutron mass was to direct neutrons (all of the same velocity) at two different targets (*A* and *B*, say). The neutrons struck the target nuclei and caused them to recoil. **(a)** If we denote the maximum recoil speeds (which occur in head-on collisions) by v_A and v_B , prove that

$$m_n = \frac{m_B v_B - m_A v_A}{v_A - v_B}$$

(use nonrelativistic mechanics). **(b)** Chadwick used hydrogen and nitrogen for his targets, and his measured speeds were $v_H = 3.3 \times 10^7 \text{ m/s}$ and $v_N = 4.7 \times 10^6 \text{ m/s}$. What was his value for m_n based on these data?

SECTION 17.7 (Nuclear Fission)

17.47 • Use the masses given in Appendix D to find the energy released in the fusion reaction (17.45) and the fission (17.46).

17.48 • One can get a rough estimate of the energy released in fission by considering Fig. 17.14, which shows the potential energy of the two fragments. The energy released is roughly the potential energy at $x = x_0$, which, in turn, is roughly $kq_1 q_2 / x_0$. (Both of these approximations will actually be overestimates.) Taking $x_0 \sim 12 \text{ fm}$, estimate the energy released in the fission $^{236}\text{U} \rightarrow ^{92}\text{Kr} + ^{144}\text{Ba}$.

17.49 • Consider the fission $^{236}\text{U} \rightarrow ^{92}\text{Kr} + ^{144}\text{Ba}$. **(a)** Using the data in Appendix D, find the number of neutrons by which the fragment ^{92}Kr exceeds its nearest stable isotope. **(b)** Do the same for ^{144}Ba . **(c)** If each fragment threw off two neutrons during the process of fission, would it then be stable? If not, what kind of radioactivity would you expect it to show? (Explain)

17.50 •• A large electrical generating plant produces about 1 GW of power. (1 GW = 10^9 W .) **(a)** What mass of uranium would a 1-GW nuclear plant consume in one year? (Assume that each fission of ^{235}U produces about 200 MeV, that the conversion of this energy to electrical energy is about 30% efficient, and that the fuel is natural uranium — of which 99.3% is ^{238}U , which produces no energy.) **(b)** What mass of coal would a 1-GW coal-burning plant consume in a year? (1 kg of coal produces roughly 33 MJ of energy, and since coal-burning plants can operate at higher temperatures, their efficiency is about 40%.) **(c)** A modern railroad car can carry about 100 tonnes. (1 tonne = 1000 kg). How many carloads of coal does the coal plant use in a year? **(d)** How many carloads of uranium does the nuclear plant use?

17.51 •• To improve the chance that the neutrons produced in fission will themselves induce further fissions, it is important to slow them down. A *moderator* is a substance with which neutrons can collide and lose energy. It must be chosen so that the neutrons lose as much energy as possible (but are not actually captured). In practice the best process for reducing the neutrons' energy is an elastic collision. Consider a head-on elastic collision between a neutron (mass m) and a nucleus

(mass M). **(a)** Use conservation of energy and momentum (nonrelativistic) to derive an expression for the neutron's fractional loss of energy, $\Delta K/K$, as a function of the mass ratio $\mu = m/M$. **(b)** Show that the loss is maximum when the masses are equal ($\mu = 1$). **(c)** This result suggests that hydrogen — in water, for example — would be the best moderator. Unluckily, ordinary hydrogen, ^1H , has the disadvantage that it can capture neutrons readily to form deuterium. Two practical alternatives with masses at least fairly close to the neutron mass are deuterium (in “heavy” water) and carbon (in the form of graphite). What is $\Delta K/K$ for these two moderators?

17.52 ••• To get a rough feeling for critical sizes, consider the following: A stream of 1-MeV neutrons impinges normally on a slab of ^{235}U , 1 cm thick. (1 MeV is the average energy of neutrons ejected in fission.) **(a)** Given that the fission cross section at 1 MeV is 1.25 barns, find the fraction, N/N_{inc} , of incident neutrons that induce a fission as they pass through the slab. [Hint: In Eq. (17.44.) we wrote the number of reactions as $N = \sigma n_{\text{inc}} N_{\text{tar}}$, where n_{inc} is the incident flux of projectiles (number/area) and N_{tar} is the total number of target particles. Here it is more convenient to use the equivalent expression

$$N = \sigma N_{\text{inc}} n_{\text{tar}}$$

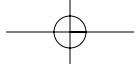
where N_{inc} is the total number of neutrons and n_{tar} is the number of ^{235}U atoms per unit area of the slab. To find n_{tar} , you need to know that the density of uranium is about 19 g/cm^3] **(b)** On average, the number of neutrons produced in a fission is $5/2$; therefore, in a self-sustaining reaction at least $2/5$ of the neutrons produced must induce further fission. Use your answer to estimate how many centimeters of slab are needed to give $N/N_{\text{inc}} = 2/5$. This answer gives the order of magnitude of the critical radius of a solid sphere of ^{235}U . **(c)** Using this estimate of the critical radius, find the approximate value of the critical mass of a sphere of ^{235}U .

SECTION 17.8 (Nuclear Fusion)

17.53 •• The principal source of energy in stars like the sun is the proton–proton cycle, the net effect of which is to fuse four protons into ^4He (plus two positrons and two neutrinos). In hotter stars the same effect is produced by a different cycle, called the **carbon**, or **CNO**, cycle. This cycle involves six reactions, parts of which are as follows:

1. $^1\text{H} + ^{12}\text{C} \rightarrow \underline{\quad} + \gamma$
2. $\underline{\quad} \rightarrow ^{13}\text{C} + e^+ + \nu$
3. $^1\text{H} + ^{13}\text{C} \rightarrow \underline{\quad} + \gamma$
4. $^1\text{H} + \underline{\quad} \rightarrow ^{15}\text{O} + \gamma$
5. $^{15}\text{O} \rightarrow \underline{\quad} + e^+ + \nu$
6. $^1\text{H} + \underline{\quad} \rightarrow ^{12}\text{C} + ^4\text{He}$

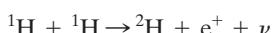
(a) Fill in the blanks in these reactions. **(b)** Verify that the net effect of one complete cycle is to replace four



94 Chapter 17 • Radioactivity and Nuclear Reactions

protons by a ${}^4\text{He}$ nucleus plus two positrons, two neutrinos, and some photons. (c) Obviously, the possibility of the cycle depends on the star's having some ${}^{12}\text{C}$ to act as a "catalyst." In addition, it requires higher temperatures than are needed for the proton-proton cycle. Explain why. Very roughly, how much higher would you expect the required temperature to be? (Remember that the high temperatures are needed to overcome the Coulomb repulsion.)

- 17.54 ••** The sun's energy comes from the proton-proton cycle, which begins with the reaction (17.61).



This reaction is exceedingly improbable, since it requires conversion of a proton to a neutron by β^+ decay during the brief collision of the two protons. Estimate the probability that a proton-proton collision will lead to this reaction, as follows: Estimate the time that the first proton (with about 70 keV) is within range of the second (time to traverse about 10 fm, say). Take the mean life for a β decay of this kind to be of order 1 minute, and find the probability that the decay will occur while the protons are within range. [Hint: Recall Eq. (17.7) and the definition of r .]

SECTION 17.9 (Nuclear Weapons)

- 17.55 •** Assuming that a certain uranium fission bomb contains 1 kg of pure ${}^{235}\text{U}$, that each fission releases about 200 MeV of energy, and that every uranium atom does fission, what would be the yield of this bomb in kilotons of TNT?

- 17.56 •** The "Little Boy" fission bomb had a yield of about 15 kilotons of TNT. Assuming that each fission of ${}^{235}\text{U}$ releases about 200 MeV of energy, what is the minimum mass of ${}^{235}\text{U}$ the Little Boy could have contained?

- 17.57 •** In a fusion bomb the two reactions (17.59) and (17.60) consume altogether three deuterons (${}^2\text{H}$ nuclei) and produce 21.6 MeV. What minimum mass of deuterium was needed to produce the 10 megatons of the first successful thermonuclear explosion by these combined processes?

SECTION 17.10 (The Theory of Alpha Decay)

- 17.58 •** Verify the values (17.74) and (17.75) for the two constants a and b that appear in the theory of alpha decay.

- 17.59 ••** One of the longest-lived alpha emitters is ${}^{232}\text{Th}$, with a half-life $t_{1/2}({}^{232}\text{Th}) = 14$ billion years, and one of the shortest is ${}^{216}\text{Ra}$ with $t_{1/2}({}^{216}\text{Ra}) = 0.18\ \mu\text{s}$. (a) Compute the ratio of these two measured half-lives and its natural log. (b) Use the formula (17.76) to predict this same natural log. (The energies released are given in Table 17.1 at the beginning of Section 17.10.)

- 17.60 •••** Evaluate the integral in Eq. (17.70), and verify the expression (17.72) for the escape probability of an alpha particle hitting the surface of a nucleus. [Hint: Remember that $U(x) = 2Zke^2/x$ and

$K = 2Zke^2/x_1$. These substitutions lead to an integral of the form

$$\int_{x_0}^{x_1} \sqrt{\frac{x_1}{x} - 1} dx$$

With the change of variables $x = x_1 \cos^2 \theta$, you can evaluate this integral as $x_1(\theta_0 - \sin \theta_0 \cos \theta_0)$, where $\theta_0 = \cos^{-1} \sqrt{x_0/x_1}$. Because $x_0 \ll x_1$, you can approximate θ_0 as $\pi/2 - \sqrt{x_0/x_1}$ and $\sin \theta_0$ as 1. Finally, you can use the relation $K = 2Zke^2/x_1$ to eliminate x_1 .]

- 17.61 •••** Occasionally, some radioactive nuclei emit other than an alpha or beta particle. In particular, it has been found that several nuclei, including ${}^{223}\text{Ra}$, can emit a ${}^{14}\text{C}$ nucleus, although this mode is very rare (of order 1 of these decays for every 10^9 alpha decays). (a) What is the offspring nucleus when ${}^{223}\text{Ra}$ emits a ${}^{14}\text{C}$ nucleus, and what is the energy released? (b) Assuming that ${}^{14}\text{C}$ emission is analogous to alpha decay, one can calculate its probability by the methods of Section 17.10. Explain why the probability $P({}^{14}\text{C})$ for escape of a ${}^{14}\text{C}$ nucleus can be obtained from Eq. (17.72) simply by replacing $2Z$ by $6Z$ and taking m to be the mass of ${}^{14}\text{C}$. (c) Calculate $P({}^{14}\text{C})$, and find the ratio $P({}^{14}\text{C})/P(\alpha)$ for the parent ${}^{223}\text{Ra}$. [Caution: Z is the charge of the offspring. The answers are quite sensitive to the value of the nuclear radius R ; to be definite, take $R = 7.5\ \text{fm}$ in both calculations.]

SECTION 17.11 (Accelerators)

- 17.62 •** A proton is accelerated from rest by a constant potential difference of 0.8 MV applied to an accelerating tube of length 1.5 m. How long does this take?

- 17.63 ••** Protons are injected into a linear accelerator so that they have $K = 250\ \text{keV}$ as they coast through the first drift tube. The frequency of the AC voltage applied to the tubes is 50 MHz and the peak potential difference across each gap is 200 kV. Consider a proton that passes from the first to the second tube properly synchronized at the peak potential drop. (a) How long must the second tube be if the proton is to arrive at the next gap in proper synchronization with the accelerating voltage? (b) How long must the tenth tube be?

- 17.64 ••** A cyclotron with a magnetic field $B = 0.9\ \text{T}$ and a maximum orbital radius of 50 cm is used to accelerate deuterons (${}^2\text{H}$ nuclei). (a) What is the maximum energy to which the deuterons can be accelerated? (b) What should be the frequency of the AC accelerating voltage? (c) If the accelerating voltage has amplitude 40 kV, how many revolutions are needed to bring the deuterons to full energy?

SECTION 17.12 (Particle Detectors)

- 17.65 •** An alpha particle from a radioactive source is directed into a cloud chamber in a uniform magnetic field of 0.80 T. If its path is curved with radius 0.40 m, what are its momentum and energy?

17.66 •• A thin foil of uranium is placed inside an ion chamber of the sort described in Example 17.7, and the foil is irradiated with neutrons to induce fissions. When a fission occurs, the ionization produced by both fission fragments can be collected as a single pulse of charge flowing around the circuit. If the total charge of one of these pulses is 9.9×10^{-13} C, what was the combined kinetic energy of the two fission fragments? (This is the method by which the large energy released in nuclear fission was discovered.) Assume that the energy loss per ionization is the same as given in Example 17.7 for alpha particles.

SECTION 17.13 (Units of Radiation)

17.67 • (a) What is the mass of 1 Ci of ^{224}Ra ? **(b)** What is it for 1 Ci of ^{226}Ra ? [Part (b) illustrates the original definition of the curie.]

17.68 • Suppose that a 60-kg person absorbs an effective dose of 5 Sv of X-rays (which, in a single exposure, would probably be fatal). **(a)** What is the total energy absorbed by the person? **(b)** Taking the person's specific heat to be about that of water, find their resulting temperature increase. (Your answer will show that heating does not play an important role in the damaging effect of radiation.)

17.69 • Medical technicians place an ion chamber 20 cm from a radioactive source (which is in its shielded container) and measure an intensity that would produce an effective dose of 3 mSv per hour in human tissue. If safety regulations limit a worker's exposure to 50 mSv per year, at what minimum distance from the source should people normally work? (Assume that they work a 40-hour week for 50 weeks per year and that absorption of the radiation in air is negligible.)

17.70 •• A 25-kg child swallows 2×10^9 Bq (about 50 mCi) of a β emitter, which ejects electrons with average energy 1.2 MeV. Assuming that all the electrons are absorbed in the body, what is the child's average absorbed dose per hour? Given that 5 Gy is a lethal dose, roughly how quickly must doctors flush out the radioactivity to save the child?

COMPUTER PROBLEMS

17.71 •• (Section 17.10) Use the data in Appendix D to find the half-lives of the isotopes ^{232}U , ^{234}U , ^{236}U , and

^{238}U and the kinetic energy released in their alpha decays. Use plotting software to make a plot of $\ln t_{1/2}$ against $K^{-1/2}$ to see how well these fit a straight line as predicted by (17.78).

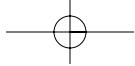
17.72 •• (Section 17.10) The half-lives of certain alpha-emitting isotopes of radium and the corresponding kinetic energies released are as follows:

A	K (MeV)	$t_{1/2}$ (s)
216	9.53	1.82×10^{-7}
218	8.55	1.4×10^{-5}
220	7.59	2.3×10^{-2}
222	6.68	38.0
224	5.79	3.16×10^5
226	4.87	5.05×10^{10}

(a) Make a plot of $\ln t_{1/2}$ against $K^{-1/2}$. **(b)** How well does the slope of your graph agree with the value predicted by Eq. (17.78)?

17.73 ••• (Section 17.3) The following data were obtained from a Geiger counter placed beside a sample of a single radioactive species. **(a)** Use suitable plotting software to make a plot of the counting rate $R(t)$ against t , including error bars. (Note: The data are the actual number of counts made in intervals of 1 minute. The uncertainty in numbers of this type is approximately the square root of the number.) Sketch an appropriate curve to fit the data, and use it to estimate the mean life τ . **(b)** Do all the same exercises using a graph of $\ln R(t)$ against t . Why is the procedure of part (b) superior?

Time (hours)	Counts in 1 min	Time (hours)	Counts in 1 min
0	99	6	15
1	78	7	15
2	64	8	7
3	36	9	3
4	21	10	6
5	22	11	4



Chapter 18

Elementary Particles

- 18.1** Elementary Particles: The Story So Far
 - 18.2** Antiparticles
 - 18.3** Pions and Muons
 - 18.4** The Four Fundamental Forces
 - 18.5** Leptons
 - 18.6** Discovery of More Hadrons
 - 18.7** The Quark Model of Hadrons
 - 18.8** The Strong Force and QCD
 - 18.9** Electroweak Interactions: The W and Z Particles
 - 18.10** Summary and Outlook
 - 18.11** High-Energy Accelerators *
- Problems for Chapter 18

*This section can be omitted without loss of continuity.

No specifiche.

GillSansTay roman
Left 9/11???

CF_EXT_FIRST

Now the smallest Particles of Matter may cohere by the strongest Attractions and compose bigger Particles of weaker Virtue; and many of these may cohere and compose bigger Particles whose Virtue is still weaker, and so on for diverse successions, until the Progression ends in the biggest Particles on which the Operations in Chymistry and the Colours of natural Bodies depend, and which by cohering compose Bodies of a sensible Magnitude.

CF_EXT_MID

There are therefore Agents in Nature able to make the Particles of Bodies stick together by very strong Attractions. And it is the Business of Experimental Philosophy to find them out.

CF_EXT_CRED

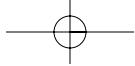
Isaac Newton, *Opticks*, 1704

18.1 Elementary Particles: The Story So Far*

“Elementary particle” is the modern name for the fundamental units of matter. As we described in Chapter 3, the concept of an ultimate, indivisible unit of matter goes back more than two thousand years to the Greek philosophers, who also gave us the name “atom,” meaning indivisible. Experimental evidence for the existence of atoms emerged gradually, beginning about 1800, and for a brief period before 1900, it was reasonable to suppose that the 90 or so different atoms were indeed the ultimate constituents of matter.

We now know that atoms are themselves divisible and so are not elementary particles in the sense that we are using the term. The first evidence that atoms can in fact be subdivided was the discovery of the electron by J. J.

*The material of this chapter depends on some of the material in Chapters 16 and 17. Specifically, you should have read Sections 16.1 through 16.3 and 17.1 through 17.4 before beginning this chapter.

Section 18.1 • Elementary Particles: The Story So Far **97**

Thomson in 1897. This was followed by Ernest Rutherford's identification of the atomic nucleus in 1910 and the discovery that nuclei are made of protons (Rutherford in 1918) and neutrons (James Chadwick in 1932). With the discovery of the neutron, it appeared that there were just three different elementary particles: the electron, proton, and neutron.

Even as the pieces of this beautifully simple picture were falling into place, evidence was accumulating that the true story was more complicated. In particular, it became clear that there were several additional particles with just as much right to be considered elementary as the electron, proton, and neutron. We have already met several of these particles in earlier chapters. For example, the positron, or antielectron, was discovered by Carl Anderson in 1932. As we have seen, this particle has the same mass and spin as the electron, but has charge $+e$. The positron has no part in the construction of normal matter, but is generally agreed to be just as elementary as its more familiar partner, the electron.

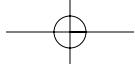
A second "extra" particle that we have already met is the neutrino, the neutral particle produced in nuclear beta decay, as proposed by Wolfgang Pauli in the early 1930s. As we have argued, the neutrino is not normally present in nuclei. Nevertheless, all of space is filled with numerous neutrinos coming from nuclear reactions in stars, and it would be difficult to exclude the neutrino from any list of elementary particles.

One other familiar particle that was already known in the 1930s is the photon. This is certainly not a *constituent* of matter in the way that the electron, proton, and neutron are; but it has a fundamental role in the structure of matter, since it is the carrier of electromagnetic forces (in a sense that we will discuss shortly). Thus, whether or not we describe the photon as an elementary particle, it is undeniably a particle with an important function in matter.

By the mid-1930s it was clear that the simple picture of a world containing just three kinds of elementary particle, the e, p, and n, was not the end of the story. In this chapter we describe the continuing quest for the truly elementary particles, whatever they may be. We will see that from the 1930s to the 1960s the picture steadily became more complicated, with the discovery of more and more particles, most of them unstable but all of them apparently elementary. Then, in the mid-1960s a new order began to appear, as evidence emerged that most of the particles that had seemed to be elementary were in fact composites, all made up from a small number of more fundamental particles called quarks (a whimsical name borrowed by the American physicist Murray Gell-Mann from James Joyce's *Finnegan's Wake*). Of the particles that were formerly considered to be elementary, only a few are still regarded as such; among these are the electron, neutrino, and photon, none of which have yet shown any evidence of an internal structure. Most of the rest, including the proton and neutron, are composites, each made up of two or three quarks.

Before we continue the story of particle physics, we should mention that this field of study is often called **high-energy physics**. The reason for this second name is easy to see: Elementary particles have sizes of 10^{-15} m and less, and certainly cannot be observed with ordinary visible light, whose wavelength is about 10^{-6} m. Instead, as we saw in our discussion of nuclear physics, these subatomic systems are probed with beams of particles, such as electrons, protons, and high-energy photons. The wavelength of these particles is given by the de Broglie relation,

$$\lambda = \frac{h}{p}$$



98 Chapter 18 • Elementary Particles

and to investigate sizes of order 10^{-15} m one needs wavelengths of this same order or less. A simple calculation using the de Broglie relation shows that this requires a kinetic energy of order 1 GeV = 10^9 eV (Problem 18.2). Some experiments today probe distances down to 10^{-17} m or less, and this requires energies of hundreds and even thousands of GeV. By atomic standards, this is an extremely high energy and amply justifies the name “high-energy physics.” To a great extent, the story of particle physics has been the story of the development of larger and larger machines to accelerate particles to higher and higher energies. We describe some of these machines in Section 18.11.

18.2 Antiparticles

In this section we review and extend the story of the positron, which proved to be the first example of what is now called an antiparticle.

The Prediction and Discovery of the Positron

Paul Dirac
(1902–1984, English)

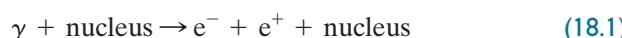


Dirac (on the left) walking with Heisenberg in the 1930s. Dirac's relativistic version of the Schrödinger equation predicted the existence of the positron (discovered a couple of years later by Anderson) and earned him the 1933 Nobel Prize in physics, which he shared with Schrödinger. He was the author of a seminal book on quantum mechanics, first published in 1930.

The existence of the positron was predicted theoretically several years before it was verified experimentally. In 1928 the English physicist Paul Dirac had proposed a relativistic generalization of the Schrödinger equation. Although Dirac's original goal was simply to find a form of the Schrödinger equation that was consistent with relativity, he accomplished much more. He found that his equation required the existence of spin, with magnitude given by the quantum number $s = \frac{1}{2}$. Therefore, if one assumed (correctly, as it turned out) that Dirac's equation described electrons, one could say that Dirac had predicted the existence of the electron's spin. Better still, the equation also predicted a magnetic moment, with the observed value (9.24).

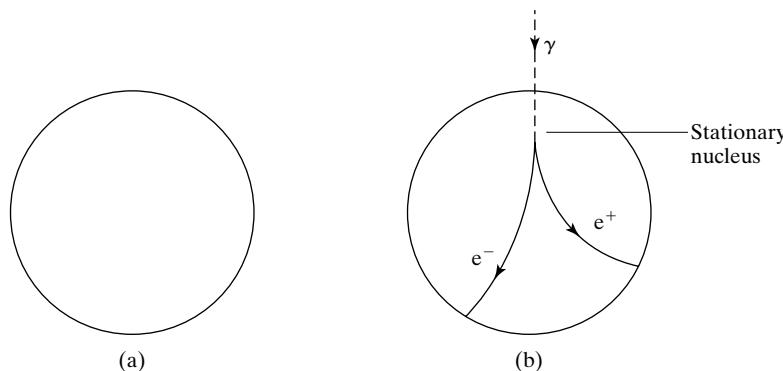
An unexpected feature of Dirac's equation, as applied to electrons, was that it required the existence of a second particle with charge +e. Since the only known particle with positive charge was the proton, Dirac thought for a while that his equation for electrons was predicting the existence of protons. However, it soon became clear that the newly predicted particle would have to have the same mass as the electron.

At the time the situation was quite confused, but in retrospect it can be easily summarized. If one accepted the existence of negatively charged electrons obeying Dirac's equation, the equation predicted the existence of a second kind of particle with the same mass and spin as the electron, but with charge +e. This particle is, of course, what we now call the positron. Furthermore, the theory predicted that an energy greater than $2m_e c^2$ should be able to create an electron–positron pair; for example, if a photon with energy greater than $2m_e c^2$ collides with a nucleus, the photon can disappear in the process of **pair production**:



where e^- and e^+ denote the electron and positron, respectively. (The need for the nucleus in this reaction is explained in Problem 18.8.) Conversely, if a positron meets an electron, they can annihilate each other with all of their rest energy (plus any kinetic energy) going into photons:



**FIGURE 18.1**

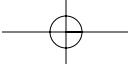
(a) An early photograph of pair production in a cloud chamber.
(b) A tracing of the same event showing the incident photon (which is invisible in the photograph). A magnetic field directed into the page caused the electron and positron to curve in opposite directions.

These predictions were unexpectedly verified in 1932 by the American experimentalist Anderson. Anderson was studying cosmic rays, the particles that bombard the earth's atmosphere from space. The primary cosmic rays are high-energy particles, mostly protons and alpha particles, with a few heavier nuclei. When these particles collide with atoms in the earth's atmosphere, they produce large numbers of secondary particles, including photons, electrons, and positrons. Anderson was studying cosmic rays using a cloud chamber, in which any moving charged particle leaves a narrow track of condensed liquid drops. These tracks can be photographed and then studied in detail. If the chamber is placed in a magnetic field, the resulting curvature of any track gives information about the particle's charge and momentum. Further, it is found that the track's thickness depends on the particle's mass, with heavier particles tending to leave thicker tracks. Thus, with care and experience, an experimenter can find most of the properties of any charged particle that passes through a cloud chamber.

Anderson had observed many tracks of electrons and protons in cosmic rays, but in 1932 he reported seeing some tracks that seemed to be made by a particle with the mass of an electron but the charge of a proton. In particular, he obtained a picture like that shown in Fig. 18.1, which is an early photograph of the pair-production process of Eq. (18.1). In the picture the incoming photon and the nucleus are invisible — the photon because it is chargeless and the nucleus because it is stationary — but the tracks of the resulting electron and positron are clearly visible. The opposite curvatures of the two tracks show that the two particles have opposite charges. That the two tracks have about equal thickness suggests (to the experienced observer) that the two particles have roughly equal masses. In the following years Anderson published conclusive evidence for the existence of a positive particle with the same mass as the electron, and there is now no doubt that his original paper was indeed the first report of what we now call the positron (named by Anderson) and that Dirac's daring predictions were correct.

The Antiproton

If the electron has an “antiparticle,” the question naturally arises: Do all particles have corresponding **antiparticles**? The answer appears to be that they do. For every particle, there is a corresponding antiparticle with the same mass and spin, and with charge of the same magnitude but opposite sign. (Notice that for a neutral particle this means that the antiparticle is also neutral.) Further, when a particle meets its antiparticle, the two can annihilate one another,



100 Chapter 18 • Elementary Particles

converting all of their rest energy into electromagnetic radiation in the form of photons. For example, a proton and antiproton can disappear in the reaction*



where \bar{p} denotes the antiproton.

Since charge is conserved and photons are uncharged, the possibility of annihilation as in (18.3) makes clear why every antiparticle must have charge opposite that of its particle. For example, the proton has charge $+e$ and, since the charge on both sides of (18.3) must balance, the antiproton must have charge $-e$.

Similar considerations apply to several other conservation laws. For example, we saw in Chapter 17 that the total number of nucleons does not change in nuclear reactions. We can reword this observation by introducing a quantity — somewhat analogous to charge — called nucleon number. We assign both the proton and neutron a nucleon number of $+1$, and other particles, like the electron and photon, a nucleon number of zero. Then we can say simply that the total nucleon number of a system is conserved in all the nuclear reactions discussed in Chapter 17. With some modifications that we discuss later, the conservation of nucleon number appears to be universally true.[†] In light of this conservation law, let us reexamine the annihilation (18.3): The proton has nucleon number $+1$, while the photon has nucleon number 0 . It then follows from (18.3) that the antiproton must have nucleon number -1 . Evidently, any particle and its antiparticle, in addition to having opposite charges, must also have opposite nucleon numbers.

Since the neutron has zero charge, so does the antineutron; that is, with regard to charge, there is no difference between n and \bar{n} . Nevertheless, the neutron and antineutron are different. An antineutron can annihilate a neutron:



whereas a neutron cannot:



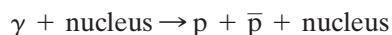
If the latter process were possible, matter as we know it would not be stable since the neutrons in nuclei would quickly disappear. Notice that in the process (18.4) the total nucleon number is conserved [$1 + (-1) = 0$], whereas in (18.5) it would not be [$1 + 1 \neq 0$]; therefore, we can say that the impossibility of (18.5) is a consequence of the law of conservation of nucleon number.

The existence of antiprotons, and antineutrons was not confirmed experimentally until 1955. Like the positron, antinucleons do not occur naturally in ordinary matter since they would be promptly annihilated. To create an

*Other reactions are also possible; for example, the p and \bar{p} can also annihilate into pions (discussed in Section 18.3). As usual in quantum physics, we cannot predict with certainty which reaction will occur — just their probabilities. In the case of the p and \bar{p} , annihilation into photons as in (18.3) is actually less probable. Nonetheless, it does occur, and this possibility can be regarded as characteristic of any particle–antiparticle pair.

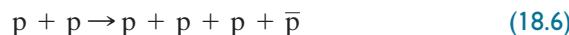
[†]The main modification is that the nucleons, n and p , turn out to be just two of a large class of particles called *baryons*, and it is the total baryon number (not just nucleon number) that is conserved.

antinucleon requires an enormous energy. For example, a proton–antiproton pair can be produced when a photon collides with a nucleus,



but the photon needs to have an energy of at least $2m_p c^2$, or roughly 2 GeV. Cosmic rays do include particles with these energies, and one might hope to discover antiprotons in cosmic rays, just as Anderson had discovered positrons. Unfortunately, 2-GeV photons are extremely rare compared to the 1-MeV photons needed to produce positrons. Thus no antiprotons had been definitely identified prior to 1955.

If antiprotons could not be found in a naturally occurring process, the only alternative was to produce them using an energetic particle that had been accelerated artificially. In 1955, Owen Chamberlain, Emilio Segrè, and coworkers at Berkeley completed an accelerator that could accelerate protons to the energy required to produce antiprotons. These protons were then fired at a target containing more protons (actually, a piece of copper, the protons being those in the copper nuclei) and produced proton–antiproton pairs in the reaction

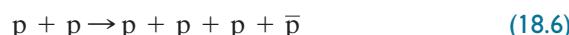


To identify the antiprotons required painstaking care since about 40,000 other particles were produced for every antiproton generated by the reaction (18.6). (In this way, the experiment was typical of many modern experiments in high-energy physics, where the events of interest must be sifted out from a background of thousands of uninteresting events.) First, a magnetic field was used to select particles with negative charge and with the momentum expected for the antiprotons. The velocity of each selected particle was then calculated from its measured time of flight between two detectors; and from the particle's momentum and velocity, one could find its mass. In this way, Chamberlain and coworkers identified several hundred particles with the mass and charge expected for the antiproton.

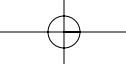
Some of these presumed antiprotons were sent through a stack of photographic films. When a charged particle passes through photographic emulsion, it leaves a trail of ionized atoms, which reveal the particle's path as a string of dark spots when the films are developed. In this way, Chamberlain and his colleagues confirmed their discovery by finding the tracks of an antiproton and the several particles produced when it annihilated with a proton.

Threshold Energy

In designing the accelerator to produce antiprotons, it was obviously essential to know how much energy was needed to allow the reaction (18.6):



At first glance one might guess that a kinetic energy $K_{\min} = 2mc^2$ (where m is the mass of the proton) would be sufficient, and this would be correct if the four final particles could be produced at rest, with total energy $4mc^2$. Unfortunately, this is not possible if the target proton is at rest: The incident proton has a large momentum. Therefore, by conservation of momentum, the four final particles must have nonzero momentum, and hence nonzero kinetic energy;



102 Chapter 18 • Elementary Particles

that is, their total energy has to be greater than $4mc^2$. Thus, the required incident kinetic energy is actually more than $2mc^2$.

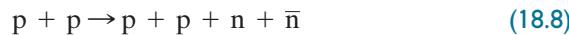
Using conservation of energy and momentum, one can show (Problem 18.10) that the minimum, or *threshold*, energy to produce the reaction (18.6) is actually

$$K_{\min} = 6mc^2 \approx 5.6 \text{ GeV} \quad (18.7)$$

This result explains why the Berkeley accelerator was designed to accelerate protons to about 6 GeV, just a little more than the threshold for production of antiprotons by the reaction (18.6). An important feature of the result is how very much energy ($6mc^2$) is needed compared to the $2mc^2$ one would at first expect. If one could arrange that the two original protons approached one another with equal and opposite momenta, the total momentum would be zero. In this case the final four particles could all be produced at rest, with a total energy of just $4mc^2$, and the threshold kinetic energy would be just $2mc^2$. This is why many modern experiments searching for extremely heavy particles are arranged with two colliding beams of particles rather than one beam incident on a stationary target. Despite the difficulties of working with two colliding beams, the saving in energy makes this arrangement worthwhile (see Problem 18.48).

Other Antiparticles

Chamberlain and Segrè were awarded the 1959 Nobel prize in physics for their discovery of the antiproton. In the same series of experiments they also found antineutrons, produced in processes such as



It is now well-established that every particle has a corresponding antiparticle. For example, the π^+ , a particle with mass $139.6 \text{ MeV}/c^2$, spin 0, and charge $+e$, has an antiparticle (denoted π^-) with exactly the same mass and spin, but with charge $-e$. Similarly, the Δ^{++} , a particle with mass $1232 \text{ MeV}/c^2$, spin $3/2$, and charge $+2e$, has an antiparticle with the same mass and spin but with charge $-2e$.

A few particles are the same as their antiparticle. For example, the photon is its own antiparticle and the same is true of the neutral pion, π^0 , which we discuss in Section 18.3. It is clear that a particle can be the same as its antiparticle only if it is electrically neutral, since particle and antiparticle must have opposite charges. Similarly, it must have nucleon number zero. For the particles that satisfy these conditions, it is often hard to be sure whether the particle and antiparticle are different. For example, it is still not completely certain that neutrinos are different from antineutrinos, although the conventional view is that they are (that is, that $\bar{\nu} \neq \nu$).

In Table 18.1 we list some of the particles discussed so far, together with their antiparticles. The list includes some examples in which particle and antiparticle are identical and several in which they are not. Notice that as you read across each line, the charge and nucleon number of each antiparticle are always equal in magnitude, but opposite in sign, to those of the corresponding particle.

Antimatter

The existence of an antiparticle for every particle implies the possibility of *antimatter*. Antimatter would be made of positrons and antinucleons, just as normal matter is made of electrons and nucleons. Indeed, several antinuclei,

TABLE 18.1

Some of the particles discussed so far, with their antiparticles. Masses are given in MeV/c^2 . In every case the mass and spin of particle and corresponding antiparticle are the same.

Name	Particle			Antiparticle			Both	
	Symbol	Charge	Nucleon number	Symbol	Charge	Nucleon number	Mass	Spin
Photon	γ	0	0	— same as particle —			0	1
Neutrino	ν	0	0	$\bar{\nu}$	0	0	$<10^{-5}$	$\frac{1}{2}$
Electron	e^-	$-e$	0	e^+	$+e$	0	0.5	$\frac{1}{2}$
Pions	$\left\{ \begin{array}{l} \pi^0 \\ \pi^+ \end{array} \right.$	0	0	— same as particle —			135	0
		$+e$	0	π^-	$-e$	0	140	0
Proton	p	$+e$	+1	\bar{p}	$-e$	-1	938	$\frac{1}{2}$
Neutron	n	0	+1	\bar{n}	0	-1	940	$\frac{1}{2}$

such as antihelium, comprising two antiprotons and two antineutrons, have been observed; and physicists at the CERN labs in Geneva have recently (2002) created large numbers of antihydrogen atoms, each comprising a positron and an antiproton. Antimatter would, of course, be extremely unstable in the presence of ordinary matter, but would be perfectly stable in isolation. It is, therefore, possible that some distant galaxies, well separated from any normal galaxies, are made entirely of antimatter, but there is at present no evidence that such galaxies exist.

18.3 Pions and Muons

In 1935 the Japanese physicist Hideki Yukawa predicted the existence of a particle with mass of order $100 \text{ MeV}/c^2$. Because this mass lay between that of the electron ($0.5 \text{ MeV}/c^2$) and that of the nucleon (about $940 \text{ MeV}/c^2$), Yukawa's proposed particle came to be called the **meson**, meaning particle of intermediate mass. When several different particles of medium mass were found, they were all referred to as mesons, and the particle that proved to be Yukawa's was named the π meson, or pion.

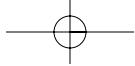
Yukawa's prediction was based on the idea that every force is, in some sense, "carried" by a corresponding particle and that the pion is required as the "carrier" of the strong nuclear force. To understand this idea, let us consider the relation of the familiar photon to electromagnetic forces. In the nineteenth century Maxwell had explained electromagnetic forces in terms of the electromagnetic fields \mathcal{E} and \mathbf{B} . As we saw in Chapter 4, it became clear at the beginning of the twentieth century that the energy in electromagnetic fields is carried by photons. Thus, in some sense at least, we can say that electromagnetic forces are carried by photons.

The exact sense in which photons carry the electromagnetic force emerges from a complete quantum theory of the electromagnetic field. This theory, called quantum-field theory, was developed in 1927 by Dirac but is beyond the scope of this book. The rough picture that emerges from quantum field theory is that the electromagnetic forces between two charged particles

Hideki Yukawa
(1907–1981, Japanese)



In 1935 Yukawa predicted the existence of the particle now called the pion, arguing that the strong nuclear force must be "carried" by a pion just as the electromagnetic force is "carried" by the photon. In 1947 the pion was discovered, with almost exactly the mass predicted by Yukawa. For this prediction he was awarded the 1949 Nobel Prize for physics.



104 Chapter 18 • Elementary Particles

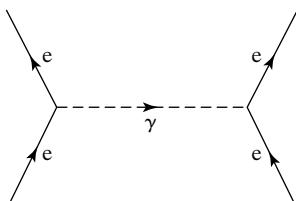


FIGURE 18.2

A Feynman diagram representing a process in which the electron on the left emits a photon, which is then absorbed by the electron on the right.

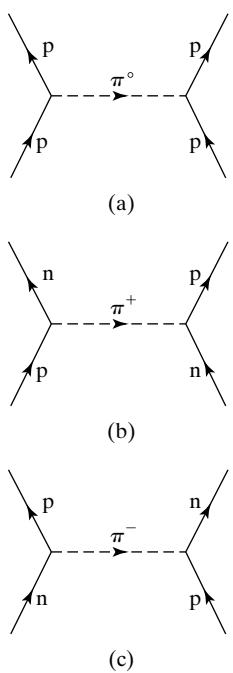


FIGURE 18.3

Three Feynman diagrams representing processes in which two nucleons exchange a pion. Two more possibilities, not shown, are that a proton and neutron can exchange a π^0 and two neutrons can exchange a π^0 .

result from one particle emitting a photon that is then absorbed by the other. This exchange of a photon between two electrons or any other charged particles can be illustrated by a **Feynman diagram** as in Fig. 18.2. These diagrams are named for the American physicist Richard Feynman, who won the 1965 Nobel prize in physics for his contributions to quantum-field theory. As suggested in Fig. 18.2, the emission of a photon by the left-hand electron changes the electron's momentum; the subsequent absorption of the photon by the right-hand electron changes its momentum by an equal and opposite amount. This exchange of momentum has the effect of a classical force between the two particles.*

Yukawa argued that if electromagnetic forces are carried by a particle (the photon), the same could be true of the strong nuclear forces. That is, the strong force between two nucleons might be the result of the nucleons exchanging some other particle, much as the photon is exchanged in Fig. 18.2. To explain the known properties of the nuclear force between all possible pairs of nucleons (p-p, p-n, n-n) it was found that there must be three kinds of exchanged particle. We now call all three of these particles **pions** and denote them π^+ , π^0 , π^- , indicating their charges of +e, 0, and -e. This means that there are several different processes involving exchange of a pion between two nucleons. Figure 18.3 shows three possibilities. In (a) two protons exchange a π^0 ; in (b) a proton and neutron exchange a π^+ ; and in (c) a neutron and proton exchange a π^- . Notice that in the last two processes the emission or absorption of a charged pion changes a proton into a neutron, and vice versa. These two *charge exchange* processes make important contributions to the resultant force between a neutron and proton.

The Mass of the Pion

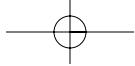
There is a relation between the mass of any force-carrying particle and the range of the corresponding force. In the case of the pion and the strong nuclear force, this relation is confirmed experimentally and lends strong support to Yukawa's theory. The connection between mass and range originates in the time-energy uncertainty relation and illustrates the essentially quantum nature of the exchange processes of Fig. 18.3. We first note that a process in which an isolated nucleon emits a pion,

$$N \rightarrow N + \pi \quad (18.9)$$

violates conservation of energy. To see this, we have only to consider the rest frame of the original nucleon. In this frame the energy of the initial nucleon is just

$$E_i = m_N c^2$$

*However, it should be emphasized that Feynman diagrams, interpreted in this way, are classical representations of an essentially quantum process. They ignore many quantum effects (such as the uncertainty principle) and can lead to contradictions if taken too literally. For example, it is hard to see how the exchange of Fig. 18.2 can lead to an attractive force, as between opposite charges. Nevertheless, a full quantum treatment does predict attraction between unlike charges. A more accurate, though less appealing description of a Feynman diagrams is that it is a convenient representation of one of the terms in an infinite series for the probability amplitude for the process concerned.

Section 18.3 • Pions and Muons **105**

whereas the energy of the final nucleon plus pion is $m_N c^2 + m_\pi c^2 + K$ (where K is their total kinetic energy). Therefore,

$$E_f \geq m_N c^2 + m_\pi c^2 > E_i$$

and the process (18.9) would violate conservation of energy by at least $m_\pi c^2$.

How, then, do exchange processes like those shown in Fig. 18.3 occur? The answer lies in the uncertainty relation $\Delta t \Delta E \geq \hbar/2$, which implies that when a system undergoes a process lasting for a time Δt , its energy has an intrinsic uncertainty

$$\Delta E \geq \frac{\hbar}{2 \Delta t}$$

Thus a violation of energy conservation by an amount $m_\pi c^2$ would be meaningless since it could never be detected (even in principle), provided that

$$m_\pi c^2 \leq \frac{\hbar}{2 \Delta t}$$

It follows that emission of a pion is possible, as long as it is reabsorbed within a time

$$\Delta t \leq \frac{\hbar}{2m_\pi c^2} \quad (18.10)$$

The maximum speed with which the pion can travel is, of course, the speed of light c . Therefore, the maximum distance the pion can travel is $c \Delta t = \hbar/2m_\pi c$. If we identify this distance with the range R of the strong nuclear force, we conclude that

$$R \approx \frac{\hbar}{2m_\pi c} \quad (18.11)$$

Knowing that the nuclear force has range $R \approx 1$ fm, we solve (18.11) for m_π and find that

$$m_\pi \approx \frac{\hbar c}{2Rc^2} \approx \frac{200 \text{ MeV} \cdot \text{fm}}{2 \times (1 \text{ fm})c^2} = 100 \text{ MeV}/c^2 \quad (18.12)$$

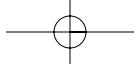
This is essentially Yukawa's original estimate for m_π . It is only a rough estimate since different assumptions about the pion's speed or about R could change the estimate by a factor of 2 or so. Thus the observed mass, $m_\pi \approx 140 \text{ MeV}/c^2$, is perfectly consistent with our estimate of $100 \text{ MeV}/c^2$. The relation (18.11) between range and mass can be applied successfully to other forces and their corresponding force-carrying particles, as discussed in Problems 18.12 through 18.14.

It is important to recognize that the exchange of a π meson between two nucleons is possible only because the pion is reabsorbed almost immediately after it is emitted. According to (18.10), the maximum time the pion can exist is of order 10^{-23} s, and there is absolutely no possibility that the exchanged pion could be observed directly. For this reason, the exchanged pions in

Richard Feynman
(1918–1988, American)



Feynman was a colorful personality who contributed to many areas of theoretical physics. He won the 1965 Nobel Prize in physics for his part in developing the quantum theory of electromagnetism (particularly the Feynman diagram), but he was equally known for his beautiful expositions of physics. His several books are gold mines of physical insight.



106 Chapter 18 • Elementary Particles

Fig. 18.3 (and, similarly, the exchanged photon in Fig. 18.2) are sometimes called **virtual particles**. As we will see shortly, however, pions *can* be observed directly in collisions where the incident kinetic energy is sufficient to supply the rest energy, $m_\pi c^2$, of the pion.

Discovery of the Muon

In 1937, Anderson (the discoverer of the positron) and his student Seth Neddermeyer discovered, among some cloud-chamber tracks of cosmic rays, a charged particle with mass around $100 \text{ MeV}/c^2$. This mass matched Yukawa's prediction almost perfectly, and the discovery had the important effect of drawing attention to Yukawa's work. Nevertheless, it became clear during the next 10 years that the particle discovered by Anderson and Neddermeyer was not Yukawa's meson. For this reason, Anderson and Neddermeyer's particles were eventually called μ particles or **muons**. (Actually, there were two such particles μ^+ and μ^- , with charges $+e$ and $-e$.) As already noted, Yukawa's particles came to be called pions, denoted π^+ , π^0 , and π^- .

In retrospect, at least, the most important difference between the pions and muons is this: As carrier of the strong nuclear force, Yukawa's pion must necessarily interact strongly with nuclei, but Anderson and Neddermeyer's muon was found to interact very weakly with nuclei. For example, it was observed that some negative muons were captured into "atomic" orbits around an atomic nucleus. If the muon were Yukawa's particle, then, once in an atomic orbit, it would be absorbed promptly into the nucleus by the strong nuclear force (Problem 18.19) and would release enough energy to disintegrate the nucleus. The orbiting muon, however, frequently remained in orbit until it decayed of its own accord; that is, it decayed in about the same time that it takes in isolation.

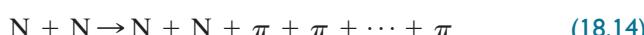
For this reason (and several more) it is quite clear that the muons discovered by Anderson and Neddermeyer were not the particles predicted by Yukawa as the carriers of the strong nuclear force. In fact, the muons seem to have no connection at all with the strong nuclear force. Rather, the μ^- and μ^+ , which are particle and antiparticle to one another, appear in almost every way to be overgrown siblings of the electron and positron, e^- and e^+ . Like the electron and positron, the muons have spin $\frac{1}{2}$ and have magnetic moments in close agreement with the predictions of Dirac's theory. Like the e^- and e^+ , muons are completely immune to the strong force.

Discovery of the Pion

By 1947, convinced that Anderson and Neddermeyer's muon was not the particle predicted by Yukawa, physicists were searching cosmic rays for signs of a second middle-mass particle that could be Yukawa's pion. If nucleons could emit and absorb virtual pions in the exchange processes of Fig. 18.3, high-energy collisions between two nucleons should be able to produce pions in processes of the form



or, given sufficient energy,



Thus, one would expect many pions to be created in the upper atmosphere when high-energy cosmic rays collide with atmospheric nuclei. On the other

hand, because pions were expected to interact strongly with atomic nuclei, the probability that they could traverse the several miles of atmosphere without being stopped by a nucleus is extremely small (Problem 18.18). Therefore, it seemed unlikely that many of the pions formed in cosmic-ray collisions would find their way down to sea level. Accordingly, a British group led by Cecil Powell began observations on the tops of mountains in the Pyrenees and the Andes. Photographic plates were left for a week or more on the mountaintops and, when developed, showed tracks of charged particles that had passed through them. After examining many hundreds of these photographs, Powell and his colleagues established that there were indeed two different kinds of medium-mass particles: the familiar muons with mass $105 \text{ MeV}/c^2$, and the long-expected pions with mass $140 \text{ MeV}/c^2$.

The first pions to be discovered were the charged π^+ and π^- . Powell and his colleagues obtained pictures in which the pions decayed into muons in the processes

$$\pi^+ \rightarrow \mu^+ + \nu \quad (18.15)$$

and

$$\pi^- \rightarrow \mu^- + \bar{\nu} \quad (18.16)$$

(The neutrinos, being neutral, cause no ionization and leave no visible tracks; nonetheless, their presence was evident from conservation of momentum.) It has since been shown that the processes (18.15) and (18.16) are the most probable decay modes of the charged pions and are the principal source of the many muons observed in cosmic rays. The lifetimes of both π^+ and π^- were found to be of order 10^{-8} s.

Powell and his colleagues found a few pictures in which a pion decayed into a muon, as in (18.15) or (18.16), and the muon then decayed by its principal decay mode, into an electron and two neutrinos:^{*}

$$\mu^\pm \rightarrow e^\pm + \nu + \bar{\nu} \quad (18.17)$$

An early picture of this sequence of events is shown in Fig. 18.4(a).

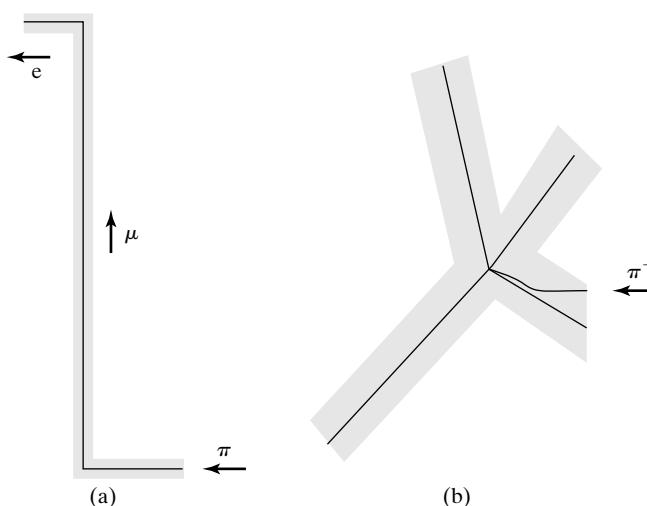
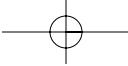


FIGURE 18.4

(a) Tracks of a cosmic-ray pion and its decay products in photographic emulsion. The pion decays into a muon and an invisible neutrino; the muon then decays into an electron and two invisible neutrinos. (b) A “star” formed when a π^- is caught in an atomic orbit and absorbed in the nucleus; the energy released blows the nucleus apart into several charged fragments.

*Two neutrinos are required in this decay because all the particles involved have spin $\frac{1}{2}$. If there were only one neutrino on the right, the final total angular momentum would have an integer quantum number, whereas the initial μ has half-integer angular momentum.



108 Chapter 18 • Elementary Particles

It was quickly verified that pions interact strongly with nuclei, as expected. For example, the negative π^- can be captured into an “atomic” orbit around a nucleus, but is then rapidly drawn into the nucleus and absorbed. The energy released breaks the nucleus apart into several fragments, as illustrated in Fig. 18.4(b).

Soon after their discovery in cosmic rays, it became possible to produce pions artificially in processes like (18.14), with protons from accelerators such as the 6-GeV machine at Berkeley. This was obviously much more convenient than waiting on mountaintops for the occasional arrival of a cosmic-ray pion, and most of the accurate measurements of pion properties were made with pions produced by accelerators.

As expected from Yukawa’s theory, it was found that in addition to the charged π^+ and π^- , there is a neutral π^0 . While the two charged pions, π^\pm , have exactly the same mass (being particle and antiparticle), the π^0 is slightly lighter:

$$m(\pi^\pm) = 139.6 \text{ MeV}/c^2$$

whereas

$$m(\pi^0) = 135.0 \text{ MeV}/c^2$$

A much more dramatic difference between the charged and neutral pions concerns their decay. The π^+ and π^- decay mainly by the processes (18.15) and (18.16), with equal mean lives of 2.6×10^{-8} s. (The mean life of any unstable particle is equal to that of its antiparticle.) The π^0 decays mainly to two photons,

$$\pi^0 \rightarrow \gamma + \gamma \quad (18.18)$$

with the much shorter mean life

$$\tau(\pi^0) = 8.3 \times 10^{-17} \text{ s} \quad (18.19)$$

The reason for the enormous difference between the lifetimes of the charged and neutral pions is that the decay (18.18) of the π^0 is induced by electromagnetic forces, while the decay of the charged pions depends on the much feebler weak force, which we discuss in the next section.

18.4 The Four Fundamental Forces

After the discovery of the muon and pion, many more particles were found, and it slowly became clear that the growing family of elementary particles could be classified into smaller groups according to their relationship with the fundamental forces of nature. In this section we discuss these fundamental forces.

By the mid-1930s it was evident that in addition to the familiar electromagnetic and gravitational forces, there had to be two other fundamental forces. First, there was the strong nuclear force that binds nucleons together in nuclei. Second, it was found that the phenomenon of β decay could not be explained in terms of the strong, electromagnetic, or gravitational forces. Therefore, there had to be a fourth force that caused β decays, and this was called the **weak force** because β decays occur so slowly compared to processes involving the strong force.

Section 18.4 • The Four Fundamental Forces **109**

The ranges and strengths of the four forces are widely different. Both gravity and electromagnetism obey inverse-square laws and can, in principle, be felt at any distance. In this sense, we say that these forces have infinite range. By contrast, the range of the strong force is only about 10^{-15} m, and that of the weak force is even smaller, being of order 10^{-18} m. At macroscopic distances, therefore, the strong and weak forces are absolutely negligible.

Even when two particles are very close, it is not necessarily the case that all four forces act between them. For example, no matter how close we bring two electrons, there is no strong force between them. However, when all four forces do act, the strong force is the most important, followed, in order of decreasing strength, by the electromagnetic, weak, and gravitational forces. In particular, the effects of gravity on individual particles are so small that they can usually be ignored entirely.

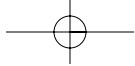
In discussing the fundamental forces, physicists tend to use the word “**interaction**” rather than “force.” The reason for this preference is that force has the connotation of a simple push or pull. This connotation is entirely appropriate when we speak of the electromagnetic force that holds an electron in an atomic orbit. However, the same basic phenomenon — electromagnetism — is responsible for processes like the creation of a photon from the energy released in an atomic transition; and in this context, it seems inappropriate to speak of the electromagnetic *force*. Today, particle physicists use “electromagnetic interaction” to describe all manifestations of electromagnetism, from the simple force between two charges to the production of photons in atomic and nuclear transitions. In the same way, “strong interaction” is used to describe the force between two nucleons, but also processes like the production of pions in energetic collisions between protons. Similarly, “weak interaction” describes the mechanism that produces the electron and neutrino in nuclear β decays.

We have already seen that a few particles, including the electron, muon, and neutrino, are not affected by the strong interactions. These particles are called **leptons** (from the Greek for “light,” since the first leptons discovered were the lightest known particles). As we describe in Section 18.5, we know of just six leptons and, as far as we can tell, all six are truly elementary particles. That is, there is no evidence that the leptons are composites made up of smaller particles.

Those particles that do feel the strong interactions are called **hadrons** (from a Greek word for “strong”). The obvious examples of hadrons are the proton, neutron, and pion, but we now know of several hundred others. As we describe in Section 18.7, it now appears that none of these particles is truly elementary. Rather, it seems that they are all built up from just six different particles called quarks, and that these six quarks are structureless elementary particles like the six leptons.

Finally, in addition to the leptons and hadrons, there are the particles that carry the forces (or, more generally, interactions). These force carriers are called **gauge particles**,* and we now believe that every force is mediated by corresponding gauge particles. For example, the gauge particles for the electromagnetic interaction are photons. Those for the strong force between quarks are called **gluons**, and those for the weak interaction are called W and Z particles. We describe gluons in Section 18.8 and the W and Z in Section 18.9.

*This use of the word “gauge” originates in classical electromagnetic theory, where it is used in connection with the measurability of the electromagnetic potentials.



110 Chapter 18 • Elementary Particles

TABLE 18.2

Classification of the several hundred observed subatomic particles into three main groups.

Name of group	Characteristic	Examples	Current status
Gauge particles	Force carriers	Photon	Elementary
Leptons	Immune to strong force	e, μ, ν	Elementary
Hadrons	Subject to strong force	p, n, π	Composite (made of quarks)

Our classification of the several hundred known subatomic particles into three main groups is summarized in Table 18.2. The last column in that table indicates that according to current evidence, the gauge particles and leptons are truly elementary, whereas the hadrons are composites, each made up of two or three elementary quarks. It is, of course, possible that future experiments will show that some of the particles that now appear to be elementary are themselves composites of some even more basic entities.

18.5 Leptons

The lepton family consists of spin- $\frac{1}{2}$ particles that are immune to the strong interaction and includes the electron, muon, and neutrino. Apart from the large difference in their masses, the first two of these, e^- and μ^- , are strikingly alike. Both carry charge $-e$, and both have an antiparticle with charge $+e$. For almost every process that involves one, there is a corresponding process involving the other. For example, the usual decay mode of the π^- is to a μ^- and antineutrino,

$$\pi^- \rightarrow \mu^- + \bar{\nu} \quad (18.20)$$

but the π^- can also decay into an e^- and antineutrino*

$$\pi^- \rightarrow e^- + \bar{\nu} \quad (18.21)$$

The other lepton that we have already met is the neutrino ν (plus its antiparticle $\bar{\nu}$, of course). In 1962 it was found that there are at least two different kinds of neutrino: The neutrino produced in association with the μ , as in (18.20), is different from that produced with the e , as in (18.21). These two kinds of neutrino are now called the μ -neutrino, ν_μ , (with antiparticle $\bar{\nu}_\mu$) and the e -neutrino, ν_e (with antiparticle $\bar{\nu}_e$). Thus, the two reactions (18.20) and (18.21) should be rewritten as

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (18.22)$$

and

$$\pi^- \rightarrow e^- + \bar{\nu}_e \quad (18.23)$$

*The decay (18.21) is very rare compared to (18.20). However, this difference can be explained by the large difference in the e and p masses (see Problem 18.24).

That these two neutrinos really are different was proven by showing that the μ -neutrinos produced in (18.22) could induce certain reactions involving muons, but could not induce the corresponding electron reactions (which *can* be induced by e-neutrinos).

In 1975 it was found that there are two more leptons: a charged τ^- and a corresponding τ -neutrino ν_τ (although the latter was only observed directly in 2000). The τ^- and its antiparticle τ^+ were produced by colliding electron and positron beams, in the reaction



and both decayed in a variety of modes, all of which involved τ -neutrinos. Like the muon, the τ particle is best described as an overgrown electron, but in this case the mass difference is even more enormous — m_τ is more than 3000 times greater than m_e :

$$m_\tau = 1784 \text{ MeV}/c^2$$

Like all of the other four leptons, the τ and its neutrino both have spin $s = \frac{1}{2}$.

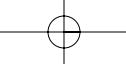
The τ particle and its τ -neutrino complete the list of known leptons. There are, as far as we know, just six leptons (each with a corresponding antiparticle). These six particles fall naturally into three pairs, or **generations**, each with one charged particle and a corresponding neutrino. This situation is summarized in Table 18.3.

TABLE 18.3

The six known leptons fall into three “generations.” The masses (in parentheses) are given in MeV/c^2 , and those for the neutrinos are upper limits. For each particle, there is a corresponding antiparticle.

Generation	Charged lepton (mass)	Neutrino (mass)
First	Electron, e^- (0.5110)	e-neutrino, ν_e ($< 10^{-5}$)
Second	Muon, μ^- (105.7)	μ -neutrino, ν_μ (< 0.17)
Third	Tau, τ^- (1784)	τ -neutrino, ν_τ (< 18.2)

Of the six particles listed in Table 18.3, only the electron has an obvious role in the structure of ordinary matter. To some extent one can understand why this is true. The electron is apparently the lightest charged particle. Therefore, there is no particle to which the electron could decay, and it *has* to be stable. By contrast, the τ and μ are both very unstable and obviously cannot be a part of normal matter. Unfortunately, we have no theory to explain why the μ and τ exist in the first place, nor why their masses are so enormous compared to m_e . It is tempting to suggest that the electron could be a bound state of some smaller objects, and that the μ and τ are excited states of this same system. However, there is at present no evidence to support this suggestion. In fact, current evidence shows that down to a scale of about 10^{-18} m , the e , μ , and τ are all structureless, elementary particles. Similarly, there is no evidence that the three neutrinos have any internal structure; and finally, no one really understands why each of the three generations includes one of these very low-mass partners.



112 Chapter 18 • Elementary Particles

18.6 Discovery of More Hadrons

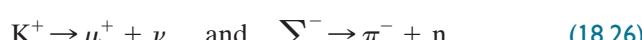
The hadron family consists of all particles that feel the strong interaction and includes the familiar proton, neutron, and pion. The advent in the early 1950s of machines that could accelerate particles to several GeV began a period of discovery of many new hadrons. In this section we describe how these particles are detected and mention some of their properties.

The Longer-Lived New Hadrons

A few of the new hadrons have mean lives of order 10^{-8} to 10^{-10} s. We will see that by the standards of particle physics this is a long time. In particular, a particle traveling near the speed of light can easily travel several centimeters in this time. If the particle is charged, one may be able to see the trail that it leaves in a suitable detector. This is illustrated in Fig. 18.5, which is a photograph made in a bubble chamber containing liquid hydrogen near its boiling point. When a charged particle traverses the chamber, tiny bubbles of hydrogen gas form near the particle's path, making a visible trail analogous to the trail of condensation left in a cloud chamber. The advantage of the bubble chamber is that the density of nuclei in the liquid is some 1000 times greater than that in the vapor of a cloud chamber; this increases the frequency of collisions by the same factor and gives the experimenter much more data in a reasonable time. In Fig. 18.5 the tracks of several negative pions are visible entering the picture from below. One of these pions has hit a stationary proton (a hydrogen nucleus) and produced two new charged hadrons, called K^+ and Σ^- , in the reaction



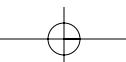
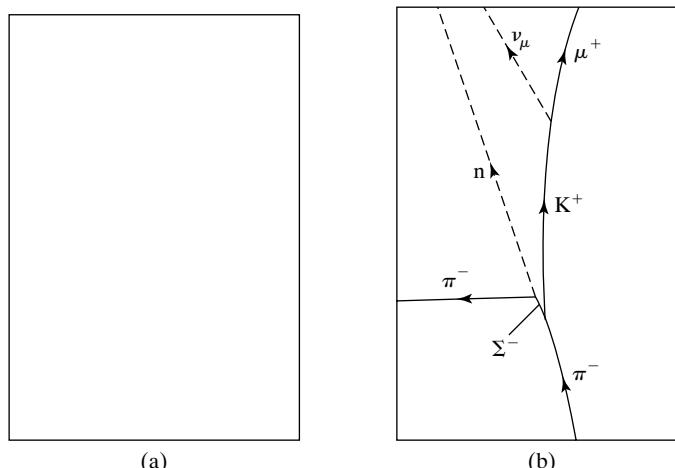
Both of the new particles produced in the reaction (18.25) have left visible trails in Fig. 18.5, and their subsequent decays can also be seen via the processes

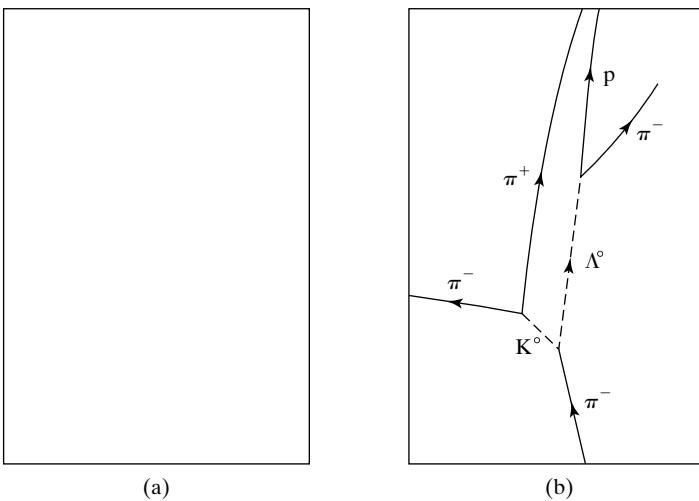


The track of each charged particle is curved because of an applied magnetic field; and from the measured curvature, one can calculate the particle's momentum. Under favorable conditions (Problem 18.35) one can also find the

FIGURE 18.5

(a) Photograph made in a hydrogen bubble chamber traversed by several negative pions. One π^- has hit a stationary hydrogen nucleus and induced the reaction $\pi^- + p \rightarrow K^+ + \Sigma^-$. (b) Tracing of the interesting tracks, including the paths of the neutral particles, which are invisible in the photograph. The Σ^- has decayed into a neutron and π^- after traveling a short distance; the K^+ has decayed into a neutrino and μ^+ shortly before leaving the picture.



Section 18.6 • Discovery of More Hadrons **113****FIGURE 18.6**

(a) Bubble-chamber photograph of the reaction $\pi^- + p \rightarrow K^0 + \Lambda^0$. Neither of the neutral particles is visible, but their subsequent decays into two charged particles are easily seen. (b) Tracing of the same sequence, including the paths of the two neutral particles.

energy and hence speed. Knowing the particle's speed and how far it traveled, one can calculate how long it lived. Then, if one has several examples to average over, one can find the mean life.

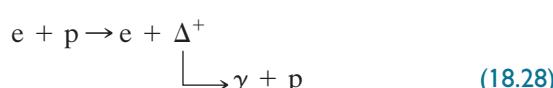
If a particle is neutral, it leaves no track in a bubble chamber or similar device. Nevertheless, its path and properties can be inferred if its decay products are charged. This is illustrated in Fig. 18.6, which shows a bubble-chamber photograph of a reaction that produces two neutral hadrons:



Neither of these neutral particles is visible, but each decays into two charged particles, producing a characteristic “vee” on the picture.

Resonance Particles

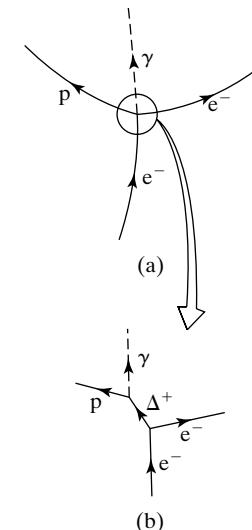
Many of the new particles decay far too quickly to leave an identifiable track like those in Figs. 18.5 and 18.6. To illustrate how these so-called resonance particles can be identified, let us consider the case of the hadron called Δ^+ . This can be produced in the reaction



where we have also indicated the subsequent decay of the Δ^+ into $\gamma + p$.* The Δ^+ lifetime is so short that it could travel only a distance of order 10^{-15} m, which is less than a nuclear diameter and is far too short to show as a visible track in any detector. Thus a photograph of the sequence of reactions (18.28) would look just like the single direct reaction

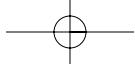


as shown in Fig. 18.7(a). Nevertheless, it is possible to identify the Δ^+ by examining many events with the overall appearance of (18.29) [some of which

**FIGURE 18.7**

(a) The reaction $e + p \rightarrow e + \Delta^+$ followed by the decay $\Delta^+ \rightarrow \gamma + p$ as seen, for example, in a bubble chamber. The Δ^+ decays too quickly to leave any track, and the observed process is indistinguishable from the single reaction $e + p \rightarrow e + \gamma + p$. (The photon would, of course, leave no visible track, but could be detected in several other ways.) (b) An imaginary enlargement by some 13 orders of magnitude shows the actual sequence of events (18.28).

*We choose this comparatively rare decay mode for convenience in the following discussion. The commonest decay modes are actually $\Delta^+ \rightarrow \pi^+ + p$ and $\Delta^+ \rightarrow \pi^+ + n$.



114 Chapter 18 • Elementary Particles

would have proceeded via (18.28), and some directly]. To understand this, we must consider the energy and momentum of the particles concerned.

The energy and momentum of the Δ^+ in (18.28) must satisfy the Pythagorean relation $E^2 = (pc)^2 + (mc^2)^2$, which implies that

$$m_\Delta c^2 = \sqrt{E_\Delta^2 - (\mathbf{p}_\Delta c)^2} \quad (18.30)$$

where m_Δ is the rest mass of the Δ^+ . Because the Δ^+ leaves no visible track, E_Δ or \mathbf{p}_Δ cannot be measured directly. However, by conservation of energy and momentum,

$$E_\Delta = E_\gamma + E_p \quad \text{and} \quad \mathbf{p}_\Delta = \mathbf{p}_\gamma + \mathbf{p}_p \quad (18.31)$$

where E_γ , E_p , and \mathbf{p}_γ , \mathbf{p}_p are the energies and momenta of the outgoing γ and proton, all of which *can* be measured. Substituting (18.31) into (18.30), we see that

$$m_\Delta c^2 = \sqrt{(E_\gamma + E_p)^2 - (\mathbf{p}_\gamma + \mathbf{p}_p)^2 c^2} \quad (18.32)$$

The relation (18.32) applies to all events involving a Δ^+ as in (18.28). However, the process (18.29) can also occur directly without formation of a Δ^+ , and such direct processes are not subject to (18.32). Given a single picture like Fig. 18.7(a), we cannot tell whether it involved a Δ^+ or not, so we cannot apply (18.32) to find m_Δ . Nevertheless, it is possible to identify the existence of the Δ^+ particle and to find its mass, provided that one has many similar pictures.

For each picture like Fig. 18.7(a), we measure the energies and momenta of the outgoing γ and proton and calculate the quantity

$$x = \sqrt{(E_\gamma + E_p)^2 - (\mathbf{p}_\gamma + \mathbf{p}_p)^2 c^2} \quad (18.33)$$

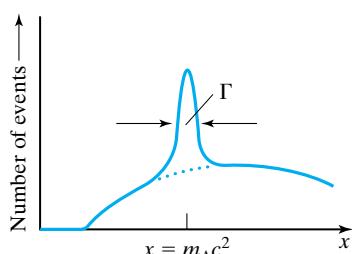


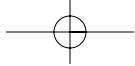
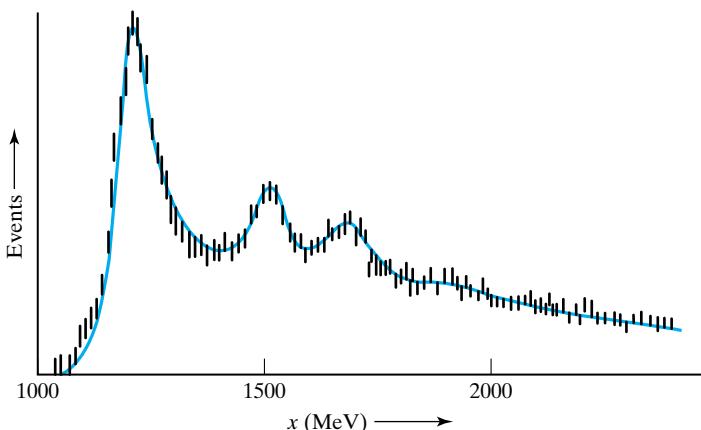
FIGURE 18.8

Schematic plot of the number of events like that in Fig. 18.7(a) against the variable x defined in Eq. (18.33). Those events in which the outgoing γ and proton came from decay of a Δ^+ all have $x = m_\Delta c^2$ and cause the spike at that value. The parameter Γ is called the width of the Δ^+ and is traditionally defined as the width of the spike at half its maximum height.

From (18.32) it is clear that x must equal $m_\Delta c^2$ for all those events in which the γ and p came from decay of a Δ^+ . On the other hand, if one calculates x for those events (18.29) that did *not* involve a Δ^+ , a continuous range of values is obtained. Therefore, if we draw a histogram, plotting number of events against the observed values of x , we should get a curve like that shown in Fig. 18.8. The smooth curve (including the dotted curve under the spike) corresponds to the direct events in which a Δ^+ was not produced; the sharp spike at $x = m_\Delta c^2$ corresponds to those events (18.28) in which the outgoing γ and proton came from decay of a Δ^+ , and for all of which $x = m_\Delta c^2$. Turning this argument around, a plot like Fig. 18.8 is evidence for the existence of an unstable particle, and the value of x at which the spike occurs tells us the particle's rest energy mc^2 . Figure 18.9 shows real data from e-p collisions. The spike at 1232 MeV corresponds to the Δ^+ particle (with $m_\Delta c^2 = 1232$ MeV), and the other two spikes correspond to two other unstable particles.

A particle whose presence has to be inferred as in Fig. 18.8 or 18.9 is called a **resonance particle**.* In practice, resonance particles are those particles

*The name is based on an analogy with resonances in *LC* circuits. If an AC voltage of frequency f is applied to an *LC* circuit, a large current I results if f is close to the resonant frequency, but I is small otherwise. Thus a graph of I against f has the same general shape as Fig. 18.8, with a sharp spike at the resonant frequency.

Section 18.6 • Discovery of More Hadrons **115****FIGURE 18.9**

A plot similar to Fig. 18.8, but based on actual data for electron–proton collisions. The first spike corresponds to the Δ^+ particle with mass $1232 \text{ MeV}/c^2$; the other two spikes correspond to two other unstable particles, called $N(1520)$ and $N(1680)$.

whose lifetimes are too short to be measured directly. Nevertheless, their lifetimes can be measured using the time–energy uncertainty relation. As we saw in Section 6.9, an unstable quantum system has a minimum uncertainty in its energy:

$$\Delta E = \frac{\hbar}{2 \Delta t} \quad (18.34)$$

where Δt is the mean life of the system — in this case the lifetime τ . For an unstable particle, ΔE is the uncertainty in the rest energy, and this uncertainty is the reason that the spike in Fig. 18.8 cannot be perfectly sharp. Particle physicists usually characterize the width of this kind of spike by the parameter Γ that is shown in Fig. 18.8 and is called the **full width at half maximum**. Since we have always defined ΔE as the half-width, it follows that $\Gamma = 2 \Delta E$, and (18.34) can be rewritten as $\Gamma = \hbar/\Delta t = \hbar/\tau$ or

$$\tau = \frac{\hbar}{\Gamma} \quad (18.35)$$

By measuring the width Γ from Fig. 18.8, we find the mean life τ .

Example 18.1

The rest energy of the Δ^+ particle is uncertain, with a measured width $\Gamma = 115 \text{ MeV}$. What is the mean life of the Δ^+ ?

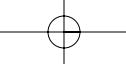
According to (18.35), the mean life τ is given by

$$\tau = \frac{\hbar}{\Gamma} = \frac{6.6 \times 10^{-16} \text{ eV} \cdot \text{s}}{115 \times 10^6 \text{ eV}} = 5.7 \times 10^{-24} \text{ s}$$

This value justifies our earlier claim that a Δ^+ particle — even if traveling at a substantial fraction of c — can cover only a distance of order 10^{-15} m .

The Strange Particles

One of the early puzzles about the many new hadrons found in the 1950s was the wide range in their lifetimes. Based on the known strength of the strong interaction, calculations showed that an unstable hadron should have a mean life



116 Chapter 18 • Elementary Particles

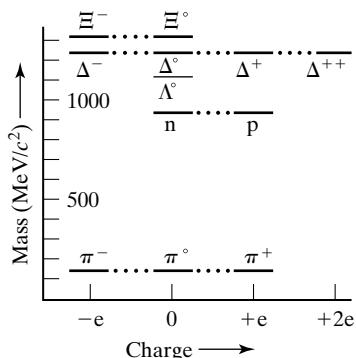


FIGURE 18.10

A plot of mass against charge for five examples of the many hadron multiplets. Except with the nucleon doublet (n and p), the same letter is used for all members of each multiplet, with the different charges shown as superscripts. On this scale, the mass differences within each multiplet cannot be seen.

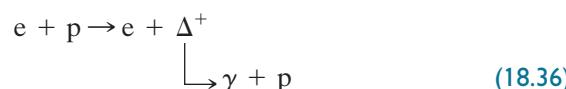
of order 10^{-23} s (like the Δ^+). Thus, the real puzzle was the enormously long lives of particles like the Σ and Λ with lifetimes of order 10^{-10} s (some 13 orders of magnitude longer than expected). For this reason, these particles were called **strange particles**. The theoretical efforts to understand their long lives contributed to the development of the quark model of hadrons, which we describe in Section 18.7.

Multiplets

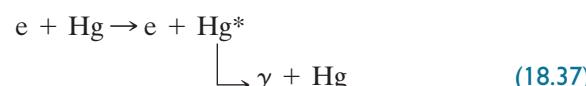
A striking feature of many of the new hadrons is that they occur in families, or **multiplets**, whose members have similar masses but different charges. For example, the Δ^+ is one of a quartet of particles, Δ^{++} , Δ^+ , Δ^0 , Δ^- , all of whose masses are close to $1232 \text{ MeV}/c^2$ but whose charges are $+2e$, $+e$, 0, and $-e$. A more familiar example is the doublet consisting of the proton and neutron, and another is the triplet of pions, π^+ , π^0 , and π^- . Several examples of multiplets, ranging from a singlet to a quartet, are shown in Fig. 18.10. Apart from the difference in their charges, the particles of a multiplet seem to be almost identical; in particular, they respond to the strong interaction in the same way. The existence of these multiplets was another stimulus for the development of the quark model of hadrons (Section 18.7).

Baryon Number

We are going to see in the next section that many of the new hadrons proved to be excited states of the neutron and proton. For example, the Δ^+ is now recognized to be an excited state of the proton. This conclusion is suggested by the process already discussed,



It is possible to interpret the first event as the excitation of the proton to a state of higher energy (called Δ^+); in the subsequent decay the excess energy creates a photon and the excited state drops back to its ground state, the proton. (Historically, the excited state was not recognized as such, and was given the separate name Δ^+ .) In this view the process (18.36) is closely analogous to the Franck–Hertz experiment (Section 5.10)



in which an electron excited a mercury atom, which subsequently emitted a photon and dropped back to its ground state. In either case the energy of the excited state (Δ^+ or Hg^*) can be deduced by observing the outgoing photon or, equivalently, by measuring the energy lost by the incident electron.

In Chapter 17 we noted that in all the processes of nuclear physics the total number of nucleons is conserved. In the light of reactions of the type $e + p \rightarrow e + \Delta^+$, this law must obviously be modified to include the excited states of nucleons. The word *nucleon* has been reserved for the neutron and proton, and the word **baryon** (from the Greek “heavy”) has been introduced to mean *either* one of the two nucleons *or* any of a family of similar particles, including excited states like the Δ^+ . (We will see in Section 18.7 that a baryon can be succinctly defined as any bound state of three quarks.) No process has

ever been observed in which the total number of baryons changes; that is, as far as we know, the baryon number, denoted B , of an isolated system is an absolutely conserved quantity.

All of the baryons, except for the neutron and proton, are unstable with very short lifetimes. Therefore, the only baryons present in normal matter are the two nucleons. In the reactions discussed in Chapter 17 there was insufficient energy to produce any of the heavier baryons; in that case, therefore, the conservation of baryon number reduced to conservation of nucleon number.

18.7 The Quark Model of Hadrons

We have said that many of the new hadrons seem to be excited states of the neutron or proton. This requires that the neutron and proton must themselves be composite systems, constructed from other, more truly elementary objects. In the early 1960s the American physicists Murray Gell-Mann and (independently) George Zweig showed that if one assumed that there are three different kinds of particles, which Gell-Mann called **quarks**, then every known hadron could be explained as a composite of two or three quarks.

The three quarks that Gell-Mann and Zweig proposed have come to be called “up” (or u), “down” (or d) — both with mass around $350 \text{ MeV}/c^2$ — and “strange” (or s) — with mass around $500 \text{ MeV}/c^2$. Each is a spin- $\frac{1}{2}$ particle, and each has a corresponding antiparticle (\bar{u} , \bar{d} , and \bar{s}). In Table 18.4 we list the principal assumed characteristics of these three quarks. (As we describe later, there are now thought to be six quarks in all.) The most startling of the properties listed in Table 18.4 is the charge of the quarks. Gell-Mann and Zweig found that they had to assume the quarks carry *fractional charge*, $2e/3$ for the u and $-e/3$ for the d and s . Since no one had ever observed fractional charges, this was a very bold assumption. Nevertheless, it has now been confirmed in several ways.

TABLE 18.4

The original three quarks proposed by Gell-Mann and Zweig.

Name	Symbol	Charge	Baryon number	Mass (MeV/c^2)	Spin
Up	u	$2e/3$	$\frac{1}{3}$	≈ 350	$\frac{1}{2}$
Down	d	$-e/3$	$\frac{1}{3}$	≈ 350	$\frac{1}{2}$
Strange	s	$-e/3$	$\frac{1}{3}$	≈ 500	$\frac{1}{2}$

Using the properties shown in Table 18.4, it is easy to understand the quark model of hadrons. All baryons (the proton, neutron, and their heavier relatives) are explained as bound states of three quarks. Each quark is assigned baryon number $B = 1/3$, which guarantees that all baryons have $B = 1$. The proton is the lowest bound state of two up quarks and one down, and the neutron is the lowest state of one up and two downs:

$$\text{proton} = (\text{uud}) \quad \text{neutron} = (\text{udd}) \quad (18.38)$$

Notice how this gives both p and n their correct charges:

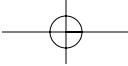
$$q_p = \left(\frac{2}{3} + \frac{2}{3} - \frac{1}{3}\right)e = e \quad q_n = \left(\frac{2}{3} - \frac{1}{3} - \frac{1}{3}\right)e = 0$$

Murray Gell-Mann

(born 1929, American)



Gell-Mann got his PhD from MIT at 22 and was a full professor at Cal Tech by the age of 26. In 1961 he predicted the existence of the omega-minus particle, which was found in 1964 (see Fig. 18.11), and in 1964 he proposed the idea that all hadrons are made up of quarks. He won the Nobel Prize in physics in 1969.



118 Chapter 18 • Elementary Particles

Since the u and d quarks have nearly the same mass, it immediately follows that the same is true of the proton and neutron. Because both the nucleons are known to have spin $\frac{1}{2}$, it is clear from (18.38) that two of the three quarks must have their spins antiparallel to give the observed total spin $s = \frac{1}{2}$.*

The Δ^+ is an excited state of the same three quarks (uud) that make up the proton. In fact, the four particles Δ^{++} , Δ^+ , Δ° , Δ^- are simply the corresponding states of the four combinations

$$\Delta^{++} = (uuu) \quad \Delta^+ = (uud) \quad \Delta^\circ = (udd) \quad \Delta^- = (ddd) \quad (18.39)$$

Since the observed spin of the Δ particles is $s = 3/2$, the three quark spins must all be parallel in this case.[†]

The strange baryons, like the Σ and Λ described in Section 18.6, are simply those baryons that include a strange quark s. For example, the particle called Σ^+ consists of two u quarks and one s:

$$\Sigma^+ = (uus)$$

The Σ^+ is the lowest state of these three quarks, but because the s quark is heavier than the d, the Σ^+ is heavier than the proton (uud). In fact, the Σ^+ is sufficiently heavy to decay into a proton and π° . However, this requires that the s quark convert into a d quark, and the only interaction that can convert one type of quark into another is the weak interaction. Thus, the process $\Sigma^+ \rightarrow p + \pi^\circ$ can occur only via the weak interaction, and this explains the “strangely” long life of the Σ^+ .

The π mesons are explained as quark–antiquark pairs; for example,

$$\pi^+ = (u\bar{d}), \quad \pi^- = (\bar{u}d)$$

In fact, the name *meson* is now used to denote any bound state of one quark and one antiquark. For example, the particles K^+ and K° mentioned in Section 18.6 are strange mesons, containing one strange quark (actually, an \bar{s}):

$$K^+ = (u\bar{s}), \quad K^\circ = (d\bar{s})$$

Notice that since all quarks have baryon number $B = 1/3$, while antiquarks have $B = -1/3$, it follows that all mesons have baryon number zero.

It is found that *all* hadrons are either baryons (three quarks) or mesons (a quark and an antiquark). One could imagine various other combinations — two quarks or two antiquarks, for example — but these seem never to occur. We shall return to this point in Section 18.8.

The quark model is much more than a mere description of how hadrons are built up from quarks. For example, the scheme predicts the existence of

AU"
will?

*Actually, the “spin” of the nucleon is the total angular momentum of the three quarks, and this could include orbital angular momentum as well as spin. However, in the lowest states the orbital angular momentum is zero and the conclusion stated is correct. As one would expect, there are excited states with nonzero orbital angular momentum as well.

[†]You may reasonably ask why in (18.38) the two nucleons do not have two more companions consisting of the combinations (uuu) and (ddd), as in (18.39). The answer (whose details are too complicated to give here) is that the Pauli principle forbids these two combinations for the $s = \frac{1}{2}$ states (but allows them for $s = \frac{3}{2}$).

Section 18.7 • The Quark Model of Hadrons **119**

multiplets of hadrons, such as the Δ quartet, which we have seen consists of corresponding spin- $\frac{3}{2}$ states of the four combinations

$$(uuu) \quad (uud) \quad (udd) \quad (ddd) \cdots \Delta(1232)$$

(The parentheses after the symbol Δ show the mass in MeV/c^2 .) Because the u and d quarks have very nearly the same mass, the same is true of these four Δ particles. Because the s quark is a little heavier than the u and d, the corresponding states containing s quarks are heavier than their nonstrange counterparts. Thus the quark model requires a multiplet of three baryons, somewhat heavier than the Δ particles, each with one s quark:

$$(uus) \quad (uds) \quad (dds) \cdots \sum(1385)$$

These have been observed and are called Σ particles. In addition, there should be two corresponding particles, each with two s quarks:

$$(uss) \quad (dss) \cdots \Xi(1530)$$

called Ξ (Greek capital letter xi), and finally, a singlet with three s quarks:

$$(sss) \cdots \Omega(1672)$$

called the Ω^- . Altogether the quark model predicts a **supermultiplet** of $(4 + 3 + 2 + 1)$ or 10 particles, all with the same wave functions and with masses that increase with the number of s quarks. The prediction of this supermultiplet was made before the Ω particle had been observed, and the subsequent discovery of the Ω , with the predicted mass and decay modes, was an important step in the development and acceptance of the quark model. A photograph made in 1964 of the production and decay of the first Ω^- to be discovered is shown in Fig. 18.11.

In Tables 18.5 to 18.7 we have listed the members of three low-mass supermultiplets, showing how all of these particles are made up from just the three quarks, u, d, and s. Tables 18.5 shows the spin- $\frac{1}{2}$ baryon octet, which includes the proton and neutron; 18.6 shows the spin- $\frac{3}{2}$ decuplet just described;

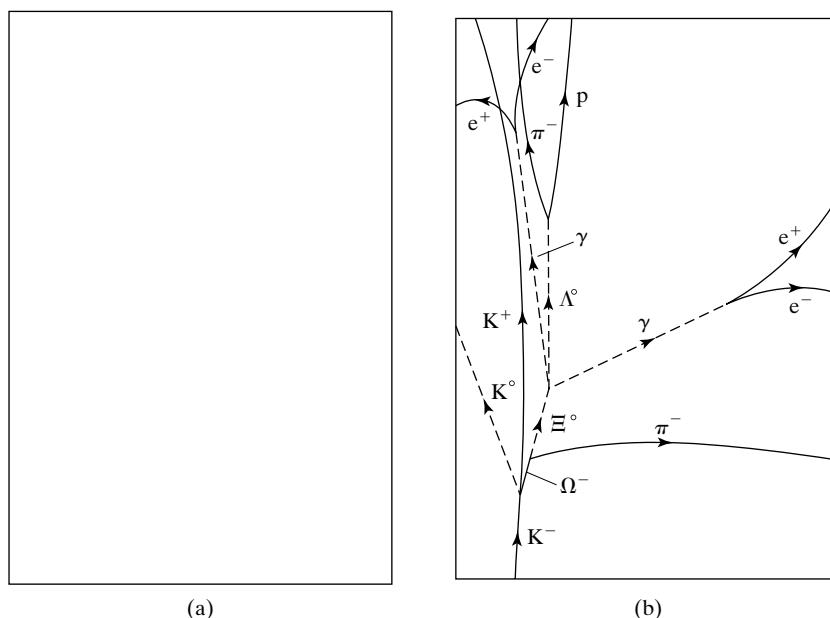
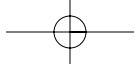


FIGURE 18.11

Photograph of the production of an Ω^- baryon in a liquid-hydrogen bubble chamber. Many K^- mesons entered the picture from below, and one produced the Ω^- in the reaction $K^- + p \rightarrow \Omega^- + K^+ + K^0$. The Ω^- traveled about an inch before decaying into a π^- and Ξ^0 . The neutral Ξ^0 is invisible, as are its decay products, γ , Λ^0 , and γ , but all of these neutrals eventually reveal themselves, the photons by producing e^+e^- pairs and the Λ^0 in the decay $\Lambda^0 \rightarrow \pi^- + p$.


120 Chapter 18 • Elementary Particles
TABLE 18.5

The lowest-mass baryon supermultiplet. The eight members of this octet are arranged in four multiplets. Each multiplet is labeled by its symbol, followed by its average mass in MeV/c^2 ; each particle symbol is followed by its actual mass. All the members of a given supermultiplet have the same spin, in this case $s = \frac{1}{2}$.

Multiplet	Particle	Quark content	Mean life (s)
N(939)	p(938.3)	uud	∞
	n(939.6)	udd	896
$\Lambda(1116)$	$\Lambda(1115.6)$	uds	2.63×10^{-10}
$\Sigma(1193)$	$\Sigma^+(1189.4)$	uus	7.99×10^{-11}
	$\Sigma^{\circ}(1192.5)$	uds	7.4×10^{-20}
	$\Sigma^-(1197.4)$	dds	1.48×10^{-10}
$\Xi(1318)$	$\Xi^{\circ}(1314.9)$	uss	2.90×10^{-10}
	$\Xi^-(1321.3)$	dss	1.64×10^{-10}

and 18.7 shows the spin-0 meson octet, which includes the three pions. These tables illustrate several general properties of the quark model. For example, the theory predicts that the number of particles in a baryon supermultiplet can be only 1 (a singlet), 8 (an octet as in Table 18.5), or 10 (a decuplet as in Table 18.6); similarly, the number of particles in a meson supermultiplet can be only 1 or 8 (as in Table 18.7). All of the several observed supermultiplets obey these

TABLE 18.6

The second lowest baryon supermultiplet is the spin- $\frac{3}{2}$ decuplet. Although the masses of the particles within each multiplet vary slightly, these differences are not shown since they are all much smaller than the widths. Similarly, the small differences in lifetimes are not included.

Multiplet	Particle	Quark content	Mean life (s)
$\Delta(1232)$	$\Delta^{++}(1232)$	uuu	6×10^{-24}
	$\Delta^+(1232)$	uud	
	$\Delta^{\circ}(1232)$	udd	
	$\Delta^-(1232)$	ddd	
$\Sigma(1385)$	$\Sigma^+(1385)$	uus	2×10^{-23}
	$\Sigma^{\circ}(1385)$	uds	
	$\Sigma^-(1385)$	dds	
$\Xi(1530)$	$\Xi^{\circ}(1530)$	uss	7×10^{-23}
	$\Xi^-(1530)$	dss	
$\Omega(1672)$	$\Omega^-(1672)$	sss	8×10^{-11}

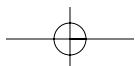


TABLE 18.7

The lowest meson octet, all of whose members have spin $s = 0$. The π° and η are mixtures of $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$. The K° and \bar{K}° each have two components with different mean lives.

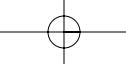
Multiplet	Particle	Quark content	Mean life (s)
$\pi(138)$	$\pi^+(139.6)$	$u\bar{d}$	2.6×10^{-8}
	$\pi^\circ(135.0)$	$u\bar{n}$ and $d\bar{d}$	8.4×10^{-17}
	$\pi^-(139.6)$	$d\bar{u}$	2.6×10^{-8}
$K(496)$	$K^+(493.6)$	$u\bar{s}$	1.2×10^{-8}
	$K^\circ(497.7)$	$d\bar{s}$	5.2×10^{-8} and 8.9×10^{-11}
$\bar{K}(496)$	$\bar{K}^\circ(497.7)$	$s\bar{d}$	5.2×10^{-8} and 8.9×10^{-11}
	$K^-(493.6)$	$s\bar{u}$	1.2×10^{-8}
$\eta(549)$	$\eta^\circ(548.8)$	$u\bar{u}$, $d\bar{d}$, and $s\bar{s}$	6.1×10^{-19}

predictions. The number of hadrons is far too large for each particle to be identified by a separate letter. Instead, as illustrated in the tables, a single letter is used to label several similar particles, and the identification is completed by giving the particle's charge as a superscript, and its mass (in MeV/c^2) in parentheses [for example, Σ^+ (1189.4) in Table 18.5 and its excited state Σ^+ (1385) in Table 18.6].

Each of the supermultiplets in Tables 18.5 to 18.7 is made up of several multiplets whose members are very close in mass. All of the particles in a single supermultiplet have many properties in common, and this is even more so for the particles within a single multiplet. For example, all members of a given supermultiplet have the same spin; the masses in a given supermultiplet are at least roughly equal, while those within a multiplet are very close. On the other hand, the lifetimes of particles within a single supermultiplet are often very different. This is because relatively small differences in mass can radically affect the possibilities for decay. For example, we see in Table 18.5 that the lifetime of the neutron is 896 s, while that of the proton is infinite (that is, the proton is stable); this is because m_n , although very close to m_p , is nevertheless greater than $m_p + m_e$, which allows the neutron to decay, whereas the proton cannot. Similar circumstances explain the anomalously long life of the Ω in Table 18.6 and the huge difference between the lives of the π^+ and π° in Table 18.7.

The 26 particles listed in Tables 18.5 to 18.7 are only a sampling of the many hadrons that have been observed. However, the majority of the known hadrons are simply excited states of those listed here, and we do not need to list them all now that we understand how they come about (any more than we needed to list all the excited states of the hydrogen atom once we understood their origin).*

*A complete listing of the known particles, called “Particle Physics Booklet,” is published every two years and can be obtained from the Particle Data Group, MS 90-308, Lawrence Berkeley Laboratory, Berkeley, CA 94720, or by visiting the Web site <http://pdg.lbl.gov/>.



122 Chapter 18 • Elementary Particles

Direct Evidence for Quarks

If hadrons are constructed from quarks, one would naturally hope to be able to break a hadron apart and observe its individual constituents. Since quarks carry fractional charges ($2e/3$ and $-e/3$), they should be easily identifiable once they are separated. However, extensive searches have produced no verifiable observations of any fractional charges. One explanation could be that the energy needed to split a hadron apart is more than the energy available in any experiment currently possible. However, this seems unlikely, since cosmic-ray particles have been observed with energies of 10^{22} eV; although energies this high are admittedly rare, one would nevertheless expect them to produce *some* free quarks, a few of which should be observed. The alternative is that although hadrons are constructed of quarks, it is actually *impossible* to break the quarks apart. This is the explanation that is currently favored. As we describe in Section 18.8, the present theory of the strong force between quarks suggests that an infinite energy is needed to pull apart either the three quarks in a baryon or the quark-antiquark pair in a meson — a phenomenon called **quark confinement**.

Although no one has observed an isolated quark, and many physicists believe no one ever will, there is compelling evidence that there are quarks inside hadrons. The first such evidence came from Stanford in California, where a 2-mile-long linear accelerator produced electrons with many GeV of energy. These electrons were fired at protons and neutrons, and the resulting reactions observed. Because the electron is a point particle and interacts mainly via the well-understood electromagnetic interaction, it is a particularly suitable probe of the structure of nucleons. The first clear conclusion of these experiments was simply that nucleons are not point particles; rather, they are extended objects with charge* spread over a distance of order 10^{-15} m.

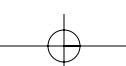
The original experiments at Stanford have been repeated at ever-increasing energies, and hence with improving spatial resolution. The results of these experiments are conclusive: The electrons scatter off the nucleon at much larger angles than would be expected if the nucleon were a continuous distribution of matter. Rather, the evidence implies that the nucleon contains several much smaller charged objects, whose individual size is 10^{-18} m or less. This conclusion is closely analogous to Rutherford's discovery of the atomic nucleus (Section 3.3), where the scattering of alpha particles at unexpectedly large angles indicated a localized concentration of charge and mass inside the atom. The distribution of scattered electrons observed in the Stanford experiments implies that there are three particles inside the nucleon, that they have spin $\frac{1}{2}$, and that they carry charges of $2e/3$ or $-e/3$. These conclusions are reinforced by experiments using other probes, such as muons. Taken together, these scattering experiments are accepted by most physicists as proof of the existence of quarks, just as the Rutherford scattering experiments were accepted as proof of the existence of the atomic nucleus.

More Quarks

We have so far discussed three types of quark, the u, d, and s. The possibility of a fourth type arose in connection with the decay of the meson called K° . It was expected that the K° should be able to decay into two muons,

$$K^{\circ} \rightarrow \mu^+ + \mu^- \quad (18.40)$$

*Of course, the neutron has zero total charge. Nevertheless, it is found that there are charges (equal quantities of positive and negative) distributed inside the neutron.



but extensive searches had found absolutely no sign of this decay. In 1970 it was shown (by an argument that is beyond our scope here) that if one postulated a fourth quark, with mass two to four times that of the s quark, the decay (18.40) would be unobservably improbable — thereby explaining the failure to observe it. Whether such a quark, called c for “charm,” actually existed was entirely unknown, but in 1974 the c quark was discovered in the form of a new meson that could be explained only as a bound state of a new quark (the c) and its antiparticle (\bar{c}). This new meson is called the ψ particle (or J/ψ in tribute to the two groups that discovered it, one naming it J and the other ψ):

$$\psi = (c\bar{c})$$

The mass of the ψ was found to be $3.1 \text{ GeV}/c^2$ and that of the c quark about* $1.5 \text{ GeV}/c^2$, comfortably inside the range predicted. A resonance plot from one of the experiments that discovered the ψ particle is shown in Fig. 18.12.

A few years later, yet a fifth quark, the b or “bottom,” was found in a similar way: A new meson, the Y (Greek capital letter upsilon), was discovered and was shown to be a bound state of the fifth quark b and its antiparticle \bar{b} :

$$Y = (b\bar{b})$$

The b quark is much heavier than its four predecessors, with $m_b \approx 5.0 \text{ GeV}/c^2$. Finally, a sixth quark, the “top” or t quark, was discovered in 1995 with an amazingly high mass, $m_t \approx 175 \text{ GeV}/c^2$ — heavier than most atoms!

Surprisingly enough, the discovery of six different kinds of quark[†] actually simplifies our picture of the elementary particles. Since there are six leptons, there is now a pleasing parallel between quarks and leptons. This parallel is sharpened when we realize that much like the six leptons, the six quarks fall naturally into three “generations.” The justification for this grouping, which is shown

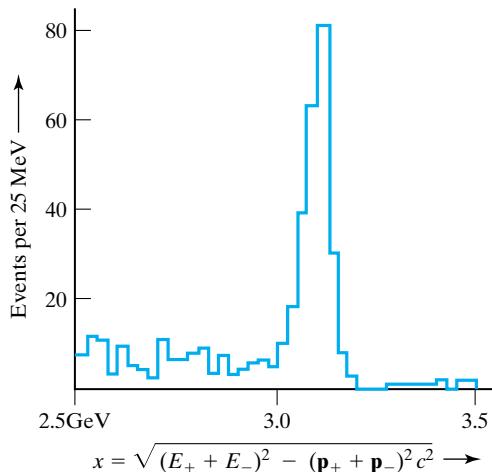
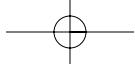


FIGURE 18.12

Data from an experiment in which the ψ particle was created in the process $p + p \rightarrow \psi + \text{other particles}$, and promptly decayed into an e^+e^- pair: $\psi \rightarrow e^+ + e^-$. This is a beautiful example of a resonance plot of the type discussed in Section 18.6 and illustrated in Fig. 18.8. The variable x is calculated from the e^+ and e^- energies and momenta; it is equal to $m_\psi c^2$ for all events that involved a ψ , but is continuously distributed for those that did not. One sees clearly that $m_\psi c^2 = 3.1 \text{ GeV}$.

*Because no quarks have been observed in isolation, their masses must be inferred from the masses of their observed bound states (such as the ψ). The mass of a bound state depends on the masses of its constituents *and* their kinetic and potential energies. Since the kinetic and potential energies are not known accurately, the same is true of the quark masses. For this reason, all quark masses are very approximate.

[†] Particle physicists use the word “flavor” for “kind of quark”; thus we speak of the six flavors u, d, s, c, b, t.



124 Chapter 18 • Elementary Particles

TABLE 18.8

The six leptons and six quarks, showing their groupings into three generations. Notice that all the particles in each column have the same charge, as indicated on the bottom line.

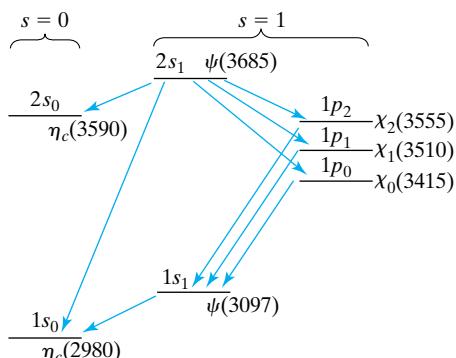
Generation	Leptons		Quarks	
First	e^-	ν_e	u	d
Second	μ^-	ν_μ	c	s
Third	τ^-	ν_τ	t	b
Charge	$-e$	0	$\frac{2e}{3}$	$-\frac{e}{3}$

in Table 18.8, is that when a quark of one type changes to a second type (which can occur only via the weak interaction), this change occurs predominantly within one generation. That is, a u changes to a d, or vice versa, much more readily than a u changes to an s, and so on. Further, just as with the leptons, it is only the first generation of quarks that plays an obvious role in normal stable matter.

Because the new quarks (c, b, t) are so much more massive than the first three (u, d, s), it turns out to be much easier to calculate the energies of their bound states. A light quark (u, d, or s) inside a hadron has kinetic energy larger than its rest energy and moves relativistically. For this reason, it has, to date, proved impossible to calculate the allowed energies of hadrons containing light quarks. By contrast, the kinetic energy of a heavy quark inside a hadron is much less than the rest energy, and the motion is nonrelativistic. This lets particle physicists use the familiar (nonrelativistic) Schrödinger equation to calculate the energies of those hadrons containing only heavy quarks. For example, using the theory of strong forces sketched in Section 18.8, one can write down the potential energy of the c and \bar{c} that comprise the ψ meson and its various excited states. Using this potential-energy function, the Schrödinger equation can be solved to give the allowed energies, and hence masses, of the ($c\bar{c}$) system. The resulting masses agree remarkably with the observed masses shown in Fig. 18.13. The states are labeled by the same scheme as used in nuclear physics, with three quantum numbers n , l , and j . Here $n = 1, 2, \dots$ for the levels of given orbital angular momentum, $l = 0, 1, \dots$ (coded as s , p , d , \dots), and j gives the total angular momentum. The three quantum numbers are given in the order nl_j (for example, $1p_2$ means $n = 1$, $l = 1$, $j = 2$). In addition, since both quarks have $s = \frac{1}{2}$, the total spin can be $s = 0$ or $s = 1$, although the $s = 0$ states are harder to observe and very few are known.

FIGURE 18.13

Energy levels of the ($c\bar{c}$) system can be calculated using the Schrödinger equation and the potential-energy function (18.41) (next section). The levels are labeled according to the scheme nl_j (for example, $2s_1$ means $n = 2$, $l = 0$, $j = 1$) and s denotes the total spin ($s = 0$ or 1). All levels shown have been observed and their experimental names are indicated, the number in parentheses being the mass in MeV/c^2 . The transitions shown are photon transitions and have all been seen.



18.8 The Strong Force and QCD

The strong force was originally introduced as the force between any two nucleons in a nucleus. However, since we now know that nucleons are bound states of quarks, it is clear that the strong force must ultimately be a force between quarks. It is natural to suppose that the strong force originates in a *strong charge* carried by quarks, in the same way that the electromagnetic force originates in the electric charge carried by charged particles. Although there is just one kind of electric charge, there appear to be three kinds of strong charge, which have come to be called **colors** — red, green, and blue. These whimsical names do not, of course, imply that quarks are colored in the ordinary sense of the word. Nonetheless, the name “color” is particularly apt, since an equal combination of all three color charges produces a neutral color charge, just as a superposition of the ordinary colors red, green, and blue can produce the neutral color white. It appears that each of the six kinds of quark (u, d, s, c, b, t) can carry any of the three different color charges. Thus any combination of three quarks can occur in an arrangement that has a total color charge of zero. In fact, all observed hadrons are color-neutral in this way, and as we will see directly, this may explain why the observed hadrons can be separated from one another whereas individual quarks cannot.

The quantum theory of the force between electric charges is called quantum electrodynamics, or QED. By analogy, the theory of the strong force between color charges is called **quantum chromodynamics**, or QCD. In QED the force-carrying particle (or gauge particle) is the massless photon. In QCD we assume that there is a corresponding massless force carrier called the **gluon**. Although there are close parallels between QED and QCD, the differences are perhaps even more important. In particular, the photon is electrically neutral, whereas the gluon carries color charge, and this leads to a striking difference between the electromagnetic and strong forces. The electric force between two electric charges falls off with distance, as indicated by the spreading lines of force in Fig. 18.14(a); by contrast, the color force between two quarks retains its strength as the quarks move apart, because the color-charged gluons attract one another and produce the lines of force shown in Fig. 18.14(b). The spreading electric lines of force describe a force $F \propto 1/r^2$ and hence a potential energy $U \propto 1/r$. The color lines of force do spread out for a short distance but then become essentially parallel; this gives a constant force and hence a potential energy that grows linearly. As a result, the potential energy of two quarks a distance r apart is thought to have the general form

$$U(r) = \alpha r - \frac{\beta}{r} \quad (18.41)$$

where α and β are constants. This is the potential that was used to calculate the levels for the $c\bar{c}$ mesons shown in Fig. 18.13. When the two quarks are close together, the potential (18.41) approximates $-\beta/r$ and resembles that of two electric charges. However, when r increases, the potential (18.41) increases without limit, so that an infinite energy would be needed to separate the quarks completely.

If these ideas are correct, no two systems with overall color charge can be separated (which explains why no isolated quark has ever been observed). On the other hand, if we consider two systems that are both color-neutral, then at large separations the color forces sum to zero and the two systems *can*

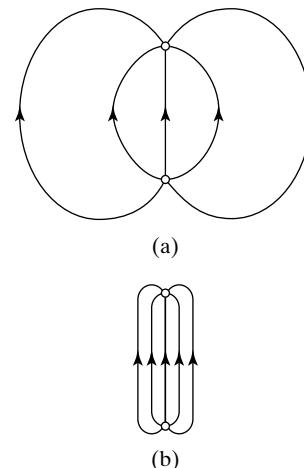
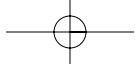


FIGURE 18.14

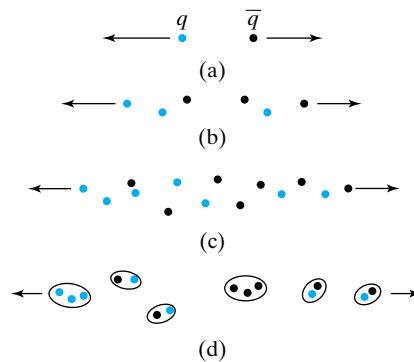
(a) The electric lines of force between two opposite electric charges. (b) Corresponding lines of force between two quarks; because the force-carrying gluons are themselves color charged, the color lines of force are drawn together and produce a constant force as the quarks move apart.



126 Chapter 18 • Elementary Particles

FIGURE 18.15

(a) A $q\bar{q}$ pair is formed in a high-energy collision and begins to separate. (b) and (c) As the q and \bar{q} move apart, energy is stored in the color field and then creates more $q\bar{q}$ pairs. (d) The many resulting quarks cluster together into color-neutral baryons and mesons, which can separate indefinitely as two “jets” of hadrons.

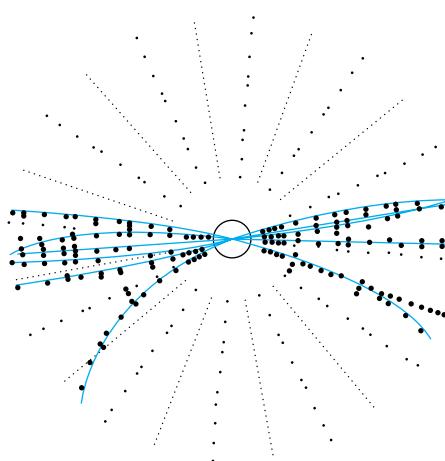


separate. The obvious examples of color-neutral systems are those comprising three differently colored quarks, for which the three colors are overall neutral, and a quark–antiquark pair, in which the color of the quark is canceled by the anticolor of the antiquark. These two examples are, of course, the observed baryons and mesons.

It is interesting to consider what would happen if one tried to pull apart two colored charges. For instance, Fig. 18.15(a) shows a q and \bar{q} flying apart after being formed in a high-energy collision. As the quarks move apart, the potential energy of the color field grows and eventually becomes so large that quark–antiquark pairs are created as in part (b) of the figure. Further separation of the original quarks produces still more pairs, as in (c), until all available energy is exhausted. At this stage the quarks can cluster together, as in part (d), to form color-neutral combinations that can separate indefinitely. Thus, what began as two outgoing quarks becomes two streams of outgoing mesons and baryons. This phenomenon is illustrated by the data from an electron–positron collision shown in Fig. 18.16, where we see two “jets” moving away from the collision region. These jets comprise the hadrons that materialized in the wakes of the q and \bar{q} that were formed when the e^- and e^+ collided. The tracks in this picture were drawn by a computer. In this respect, Fig. 18.16 is typical of modern particle experiments, in which data on the positions and momenta of the particles are recorded directly in computers; the particles’ paths can then be reconstructed and drawn by the computer to facilitate analysis by experimenters.

FIGURE 18.16

Two “jets” of hadrons produced in the wakes of a $q\bar{q}$ pair, which was formed in the high-energy collision of an e^+ and e^- at the center of the picture. The e^+ and e^- are not shown since they came in perpendicular to the picture, one from above and one from below the page. A magnetic field has curved the outgoing trajectories, allowing measurement of the momenta. The small dots show the positions of individual wires that detect the particles. The large dots show the positions at which particles were detected; this information was fed directly to a computer, which reconstructed and drew the particles’ trajectories.



Section 18.8 • The Strong Force and QCD 127

It is important to note that although the force between two color-neutral hadrons is negligible at large separations, when the hadrons are close together the individual quark-quark forces do not cancel exactly, and there is a residual net force. There is a close parallel here with the electric attraction between two atoms: When two neutral atoms are far apart, the electric force between them is completely negligible, but at distances of order the atomic size the charges inside the atoms can distribute themselves such that the attractive forces between opposite charges are greater than the repulsive forces between like charges. (As we saw in Chapter 12, this is exactly the mechanism by which neutral atoms bind together to form molecules.) In an analogous way, the different colors inside two color-neutral nucleons produce no attraction at large separations but can produce overall attraction at separations on the order of the nucleon size, and this is precisely the attractive force that binds nucleons together in a nucleus.

The attraction between two nucleons was explained by Yukawa as arising from the exchange of virtual pions, as illustrated in Fig. 18.17(a). According to QCD, this is not the ultimate explanation of the nucleon-nucleon force. Nevertheless, Yukawa's theory can be understood within the framework of QCD, as illustrated in Fig. 18.17(b), where each of the nucleons is shown as a bound state of three quarks. The three quarks in each nucleon are continually emitting and absorbing virtual gluons. (This is indicated by the shading in the picture.) When the two nucleons approach within 1 fm or so, the virtual emissions and absorptions within either nucleon can interact with those in the other. In the event shown in Fig. 18.17(b), one of the u quarks in the incoming proton has emitted a virtual gluon, which subsequently created a $d\bar{d}$ pair. The recoiling u and the newly created \bar{d} both move across to the neutron, where the d is annihilated by one of the incoming d quarks and the u is captured. The net effect is that the proton changes into a neutron and vice versa, and two quarks are exchanged, one u and one \bar{d} . Since the u and \bar{d} are precisely the constituents of a π^+ meson, we recognize that Fig. 18.17(b) is simply the quark-model version of the π^+ exchange shown in Fig. 18.17(a). The quark model of the nucleon-nucleon interaction allows many other processes besides that of Fig. 18.17(b). Nonetheless, the success of Yukawa's theory shows that those processes [like that of Fig. 18.17(b)] that are equivalent to exchange of a pion are especially important.

Although QCD successfully explains the strong interaction in terms of gluons, direct evidence for the existence of gluons is even harder to find than that for quarks. Nevertheless, the general success of QCD has convinced most particle physicists that gluons really do exist. Furthermore, there is some evidence, of which we mention two examples: The electron-nucleon scattering experiments that have shown the presence of three quarks inside the nucleon indicate that the total momentum of the three quarks is *less* than the momentum of the nucleon. It is widely believed that the missing momentum is carried by the virtual gluons that are continually bouncing back and forth among the constituent quarks.

Further indirect evidence for gluons comes from e^+e^- collisions, which sometimes produce three jets of hadrons such as those shown in Fig. 18.18. This picture is interpreted as resulting when the e^+e^- collision produces a $q\bar{q}$ pair (as in the two-jet event of Fig. 18.16), but one of the quarks promptly ejects a high-energy gluon. Like the quarks, the gluon carries color charge and cannot escape by itself. Therefore, like the two quarks, it creates a stream of $q\bar{q}$ pairs in its wake, and these emerge as a jet of hadrons. The third jet in such three-jet events is widely accepted as proof of the existence of gluons.

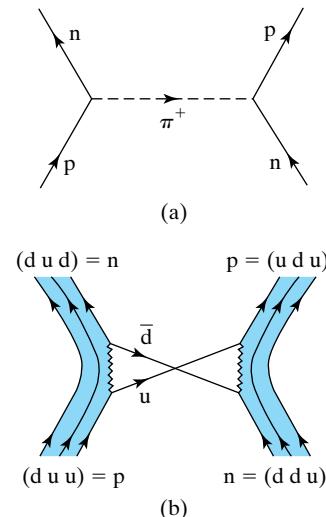


FIGURE 18.17

- (a) Yukawa explained the force between two nucleons in terms of the exchange of virtual pions.
 (b) The same process as explained in QCD. The two wavy lines represent virtual gluons.

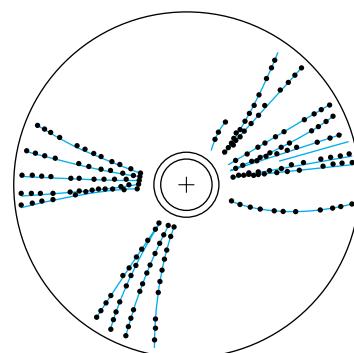
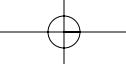


FIGURE 18.18

Three jets of hadrons produced in the head-on collision of an e^+ and e^- (which came in from directly above and below the plane of the picture). This event is similar to the two-jet event of Fig. 18.16, but here the e^+ and e^- created a q and a \bar{q} , one of which radiated a high-energy gluon. The three jets were created in the wakes of the q , the \bar{q} and the gluon.



128 Chapter 18 • Elementary Particles

18.9 Electroweak Interactions: The W and Z Particles

Perhaps the greatest triumph for particle physics in the last 30 years has been the discovery that the electromagnetic and weak interactions are different manifestations of a single phenomenon called the **electroweak interaction**. This unification has been compared to the discovery in the nineteenth century that electric and magnetic forces are two aspects of the single phenomenon that we now call electromagnetism. At first it appeared that electric and magnetic forces were independent phenomena, and their unification by Faraday and Maxwell was one of the high points of nineteenth-century physics. In a similar way, the electromagnetic and weak interactions appeared to be entirely different, and their unification was a major achievement, for which its principal authors — the Americans Sheldon Glashow and Steven Weinberg and the Pakistani Abdus Salam — were awarded the 1979 Nobel prize for physics.

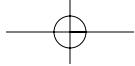
The theory of the electroweak interaction begins by postulating that the electromagnetic and weak interactions are carried by four massless gauge particles, two electrically charged and two neutral. With this assumption, there would be an almost complete symmetry between the electromagnetic and weak forces. In particular, with massless force carriers, the weak force would be a long-range force. [Remember the relation (18.11), $R \approx \hbar/(2mc)$, between the range of a force and the mass of its carrier; in particular, if $m = 0$, the force range is infinite — as is the case in electromagnetism.] Since the observed range of the weak force is very short, it is clear that this theory is not yet complete. Rather, it is postulated that the observed four force carriers must be mixtures of the original massless gauge particles, one of them still massless but three of them with mass. (The mechanism by which the originally massless particles acquire mass is called spontaneous symmetry breaking.) The surviving massless particle is the familiar photon. The three massive particles, called W^+ , W^- , and Z° , are the carriers of the weak interaction. The theory predicts their masses as

$$m(W^\pm) = 82.4 \text{ GeV}/c^2 \quad \text{and} \quad m(Z^\circ) = 93.3 \text{ GeV}/c^2 \quad (18.42)$$

both with uncertainties of about $1 \text{ GeV}/c^2$. These enormous masses account for the observed short range of the weak interaction. In fact, if we substitute the masses (18.42) in the relation $R \approx \hbar/(2mc)$, we find a range of order 10^{-18} m as observed (Problem 18.43).

Because of the large masses of the weak force carriers, the observed weak interactions are very different from the electromagnetic. Nevertheless, these differences arise in a well-understood way from a theory that is basically symmetric. At extremely high energies (thousands of GeV) the difference between the masses (18.42) and the photon mass (zero) would be negligible and the underlying electroweak symmetry should become apparent. It is to test this idea (among others) that physicists are planning to build accelerators that can produce particles with many TeV of energy ($1 \text{ TeV} = 1000 \text{ GeV} = 10^{12} \text{ eV}$).

The electroweak symmetry is important to astrophysicists interested in the birth of the universe. The present outward motion of the galaxies indicates that the universe originated in an explosion at a single point. Immediately after this “big bang” the temperatures, and hence energies, must have been high enough for the electroweak symmetry to be important and to have influenced the development of the universe we see today. Even at more mundane temperatures and energies, where the electroweak symmetry is not readily

Section 18.9 • Electroweak Interactions: The W and Z Particles **129**

apparent, the electroweak theory gives us a consistent theory of electromagnetic and weak interactions and provides an accurate method to calculate the probabilities of electroweak processes. For example, recent experiments on cesium atoms have dramatically confirmed the tiny contribution of the weak interaction to atomic transitions and verified that the predictions of the electroweak theory can be detected even at energies of only a few eV.

Discovery of the W and Z Particles

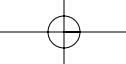
Although the success of the electroweak theory gave strong indirect evidence for the W and Z, physicists naturally hoped to observe these particles directly. Since their masses are $80\text{--}90 \text{ GeV}/c^2$, a minimum energy of order 100 GeV was obviously needed to produce them. In 1976 an accelerator called the super proton synchrotron or SPS was completed at the CERN laboratory in Europe.* This machine accelerates protons as they circulate around an evacuated circular pipe some 4 miles in circumference. The protons can be accelerated to 400 GeV, which might seem more than enough to produce the W and Z particles. Unfortunately, it is actually hopelessly inadequate because of the problem mentioned in Section 18.2: When a high-energy particle collides with a stationary particle, very little of the incident energy is available to create new particles. The reason is that conservation of momentum requires the final particles to have a large total momentum (equal to that of the original projectile). Thus the final particles have a certain minimum kinetic energy, and only the remainder of the original kinetic energy is available to create new particles. In particular, if a 400-GeV proton collides with a stationary proton, only 30 GeV of the original energy is available to create new particles (see Problem 18.48). This is totally inadequate to produce a W or Z particle.

The solution to this difficulty is obvious in theory but extremely hard to realize in practice. If one could make two high-energy particles collide head-on, in such a way that their total momentum was zero, the final particles could have zero momentum and hence zero kinetic energy, and all of the incident kinetic energy would be available to create new particles. To create two high-energy beams and arrange for them to collide head-on is obviously much harder than creating a single beam that collides with a fixed target (see Problem 18.44). Nevertheless, the SPS accelerator at CERN was modified successfully to produce two colliding beams.

One of the colliding beams was the proton beam, already available at the SPS. The second beam consisted of antiprotons. These were created in collisions of 26-GeV protons with a metal target. Such collisions produce many different particles, but suitable electric and magnetic fields could select the antiprotons and steer them to an *antiproton accumulator* — an evacuated pipe, forming a circle of diameter about 50 m, in which the antiprotons could circulate, held in orbit by a ring of magnets. Antiprotons were steadily fed into the accumulator for about a day, until there were enough for the experiment.[†] A burst of antiprotons was then guided out of the accumulator ring and boosted into the same SPS that accelerates the protons. Since the p and \bar{p} have the same mass but opposite charges, they could be accelerated simultaneously in

*CERN — Conseil Européen pour la Recherche Nucléaire — is a laboratory run jointly by 13 European countries and situated outside Geneva, Switzerland, straddling the Swiss-French border.

[†] Antiprotons were stored successfully for as long as a month — dramatically confirming the belief that when isolated from matter, antimatter is perfectly stable.



130 Chapter 18 • Elementary Particles

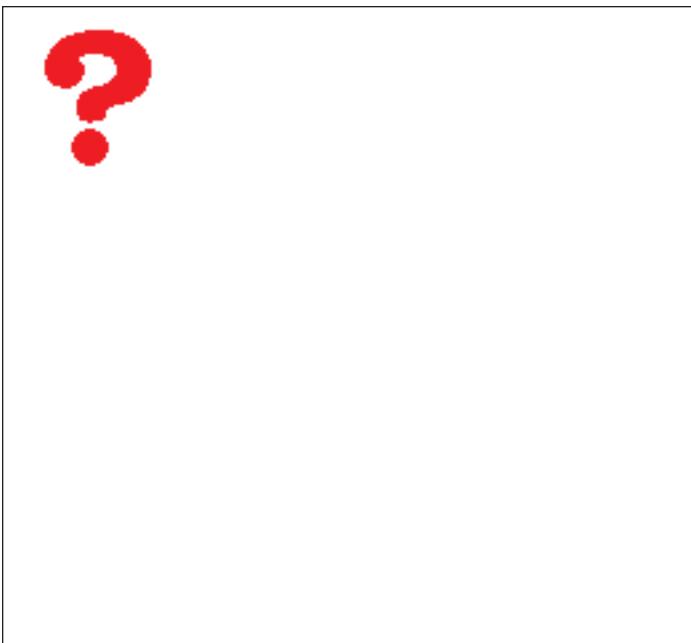


FIGURE 18.19

Part of the 2000-ton UA1 detector used to find the W^\pm and Z particles at CERN. The paths and energies of the outgoing particles were fed directly to computers for storage and later analysis.

the same machine, provided that they traveled in opposite directions. In this way the p and \bar{p} beams were brought up to 270 GeV each.

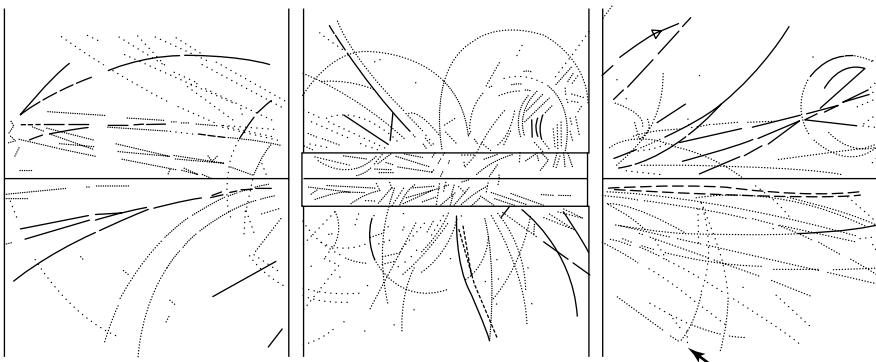
The two beams were arranged so that their paths crossed at six locations around the 4-mile circumference of the SPS tunnel. A large assembly of detectors placed all around each collision area could monitor the particles produced in the resulting $p\bar{p}$ collisions. Each detector could produce an electrical signal in response to the passage of a particle in its vicinity, and all detectors were connected directly to computers that recorded the energies of the outgoing particles and their positions and times at several points along their paths. A part of one detector assembly, which had a total mass of 2000 tons, is shown in Fig. 18.19.

The mean lives of the W and Z were expected to be of order 10^{-24} s. Therefore, they had to be identified by detecting their decay products, for example,

$$Z^0 \rightarrow e^+ + e^- \quad \text{and} \quad W^- \rightarrow e^- + \bar{\nu}_e \quad (18.43)$$

The problems of identifying these decay products were formidable indeed. First, it was calculated that only 1 collision in 10 million would produce an identifiable W or Z . Second, every collision (including those few that produce a W or Z) creates a large number of hadrons. Therefore, the CERN experiments had to observe a prodigious number of events and then, for each event, sort through an enormous clutter of hadrons to identify the hoped-for decay products of the W or Z . In addition, in the case of the W , one of the decay products was always a neutrino, as in (18.43), and the presence of this neutrino could be inferred only by measuring the energy and momentum of all other outgoing particles and comparing with the total incident energy and momentum.

Data from the collisions were taken and recorded by a large number of computers. Each event was examined briefly, and the vast majority immediately rejected as uninteresting. On average, about one event per second was selected as promising enough to be worth recording on magnetic tape — with the paths and energies of all the outgoing particles that had been detected.

**FIGURE 18.20**

Computer reconstruction of the process $p + \bar{p} \rightarrow W^- + \text{hadrons}$, followed by the decay $W^- \rightarrow e^- + \bar{\nu}_e$. The p and \bar{p} enter from left and right along the horizontal solid lines. The path of the outgoing e^- is indicated with an arrow. The path of the $\bar{\nu}_e$ had to be inferred by measurement of all the other outgoing particles.

These events could then be analyzed fully in the coming months. A few events were immediately selected as especially promising and were sent to an “express lane” for prompt processing.

After seven years of planning by nearly 200 scientists from some 20 universities and laboratories in Europe and the United States, the first full experimental run began in September 1982. After two months of almost continuous operation, the first candidate for a W particle was reported, and by the end of 1983 there were some 90 definite observations of the W and 13 of the Z . A computer reconstruction of an early example of a W^- decaying into an electron and an antineutrino is shown in Fig. 18.20. The data from these observations were enough to determine the masses of the W and Z rather accurately, and, as shown in Table 18.9, the agreement with theory was spectacular. These results were an outstanding triumph for the electroweak theory and were recognized by the award of the 1984 Nobel prize for physics to two of the leaders of the experiment, Carlo Rubbia from Italy and Simon van der Meere from Holland.

TABLE 18.9

Measured and predicted masses of the W^\pm and Z^0 in GeV/c^2 .

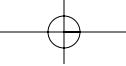
	W^\pm	Z^0
Measured	82.1 ± 1.7	93.0 ± 1.8
Predicted	82.4 ± 1.1	93.3 ± 0.9

18.10 Summary and Outlook

Let us first summarize the roster of elementary particles as sketched so far.

The Standard Model

There are 12 elementary spin- $\frac{1}{2}$ particles — six leptons and six quarks — which fall into three generations, each with two leptons and two quarks. (See Table 18.8 in Section 18.7.) The leptons, which include the familiar electron, can exist in isolation; the quarks exist only inside baryons and mesons. All 12 leptons and quarks are subject to the electroweak interactions, which are carried by four gauge particles, the massless photon, and the massive W^\pm and Z^0 , all of



132 Chapter 18 • Elementary Particles

which have been experimentally detected (by observation of their decay products). In addition, the six quarks are subject to the strong force, which is carried by the massless — and so far not directly observed — gluon.

This picture has come to be called the **standard model**, a name that reflects the confidence of most physicists that the basic facts are well-established and unlikely to change. On the other hand, we should emphasize that many questions remain unanswered: The model does not explain why the leptons and quarks have the masses observed. Why is the electron so light compared to all the other massive particles? Why are the second and third generations of particles so much heavier than the first? And indeed, why do the second and third generations occur at all?

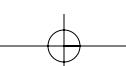
Grand Unified Theories

The electroweak theory has unified the electromagnetic and weak interactions. Many physicists believe that eventually the strong interactions, also, will be incorporated in a single **grand unified theory** (GUT). Considerable progress has been made in this direction, and some startling predictions have been made. For example, most proposed GUTs predict that the proton should be unstable with a mean life τ_p of order 10^{31} years. This life is prodigiously long, being some 21 orders of magnitude greater than the age of the universe, and for all practical purposes the proton would appear to be stable (in agreement with observation). Nevertheless, experiments have been set up that should be able to detect proton decays. So far no verifiable decays have been observed, and the experimental lower bound on τ_p is already somewhat larger than the value predicted by the simpler grand unified theories. That is, the simpler GUTs seem to be incompatible with the observed facts. Notice that because the proton is the lightest baryon, its decay — for example, $p \rightarrow \pi^+ + e^+$ — would violate conservation of baryon number. In fact, it is an essential feature of most GUTs that conservation of baryon number is violated, very weakly at least.

Another prediction of most GUTs is the existence of a magnetic monopole — a particle with a north pole but no south pole or vice versa. Such particles have never been observed, and recent searches have still not found any. Obviously, it is important to current GUTs that we establish whether the proton is unstable and whether there are magnetic monopoles. (Indeed, both of these questions are extremely interesting, regardless of their relevance to GUTs.) Meanwhile, it should be emphasized that GUTs have scored some impressive successes. For example, certain parameters of the standard model* are predicted by GUTs, and the predicted values agree well with observation.

Some physicists are already seeking a theory, even more ambitious than the GUTs, in which *all four* interactions — strong, electromagnetic, weak, and gravitational — are unified. We have emphasized that on the level of elementary particles, gravity is usually completely negligible. Nevertheless, as a matter of principle, it is desirable to have a theory that could explain *all* of the known forces. The attempts to unify gravity with the other forces go back to Einstein, who spent many years trying unsuccessfully to unify gravity and electromagnetism. Until recently the problem of incorporating gravity in a quantum theory had seemed completely intractable; but it now seems that a new approach,

*For example, the so-called Weinberg angle, which gives the degree of mixing among the massless gauge particles of the electroweak theory.



called the *theory of superstrings*,* may produce a satisfactory quantum theory of all four interactions. Nevertheless, this theory is still very speculative, and no one would claim that any answers are final.

Another speculation that deserves mention is that the leptons and quarks may themselves be composite, constructed from objects that are more truly elementary. As already noted, the occurrence of the three generations of leptons and quarks suggests this idea, and several such theories have appeared. However, there is so far no direct evidence for such a substructure, and these theories have yet to prove their worth.

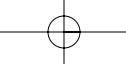
The Cosmic Connection

Particle physics, by its very nature, is concerned with the smallest scales known — currently about 10^{-18} m. At the opposite extreme is astrophysics and, especially, cosmology, the study of the structure and origin of the whole universe — whose size is of order 10^{26} m. Paradoxically, these two extreme branches have been growing increasingly dependent on one another for many years. Astronomers studying stellar spectra have needed to understand — and have contributed to — atomic physics for more than a century. A proper understanding of the energy source and evolution of stars came only after the development of nuclear physics in the 1930s. Most recently, it has become apparent that during the first moments after the big bang, the whole universe was a natural laboratory for elementary-particle physics. For the first 3 minutes the temperature was so high that no atoms or nuclei could exist, and the universe was a fiery gas of leptons and hadrons. During the first microsecond or so, the temperature may have been high enough that hadrons could not exist and there were only leptons and quarks. One of the motivations for building particle accelerators with higher and higher energies is to allow physicists to observe collisions that approximate conditions in these earliest moments of the universe.

Because the first moments of the universe presumably determined its subsequent evolution, an understanding of elementary particles has become very important to astrophysics, and recent years have witnessed several papers in astrophysics published by particle theorists, and vice versa. To conclude, we mention two examples of this kind of interplay. First, a question that has puzzled many physicists is why the universe apparently consists mainly of what we call matter (electrons and nucleons) rather than antimatter (positrons and antinucleons). It now seems conceivable that the universe was originally symmetric, with equal amounts of matter and antimatter; then, at a later time, a chance fluctuation in favor of matter could have precipitated a steady disappearance of almost all antimatter to produce the present imbalance. This scenario would require that baryon number is not conserved, as is the case in certain GUTs; thus, an improved understanding of these theories is needed before we can decide this interesting question.

Second, there is strong evidence that the mass of the universe is considerably greater than the mass of its visible components: stars, interstellar gas, quasars, and so forth. If, as the evidence now suggests, neutrinos have a small but nonzero mass, this could account for some of the invisible mass in the universe. Because these neutrinos travel near the speed of light, they are referred to as “hot” matter. It is also possible that there are other massive particles, as

*In superstring theory the elementary particles are treated as short line segments rather than points.



134 Chapter 18 • Elementary Particles

yet unobserved, that contribute to the invisible “dark matter.” These hypothetical particles would move relatively slowly and are called “cold” dark matter. Recent measurements of the gravitational clumping of galaxies suggest that the universe contains both hot and cold dark matter.

Conclusion

Obviously, extraordinary progress has been made in the last 50 years toward understanding the fundamental constituents of matter. Nevertheless, there are still plenty of questions — both theoretical and experimental — to be answered, and it is almost certain that many interesting discoveries will be made in the coming decades.

18.11 High-Energy Accelerators*

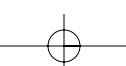
In Section 17.11 we described some of the accelerators that give charged particles kinetic energies of tens or even hundreds of MeV. These energies suffice for the study of atomic nuclei, but the study of elementary particles requires energies from several GeV to several thousand GeV. The two types of accelerator that have succeeded in producing these immense energies are the *linear accelerator*, or *linac*, which accelerates the particles along a straight path, and the *synchrotron*, in which the particles follow a circular path of fixed radius.

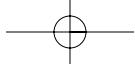
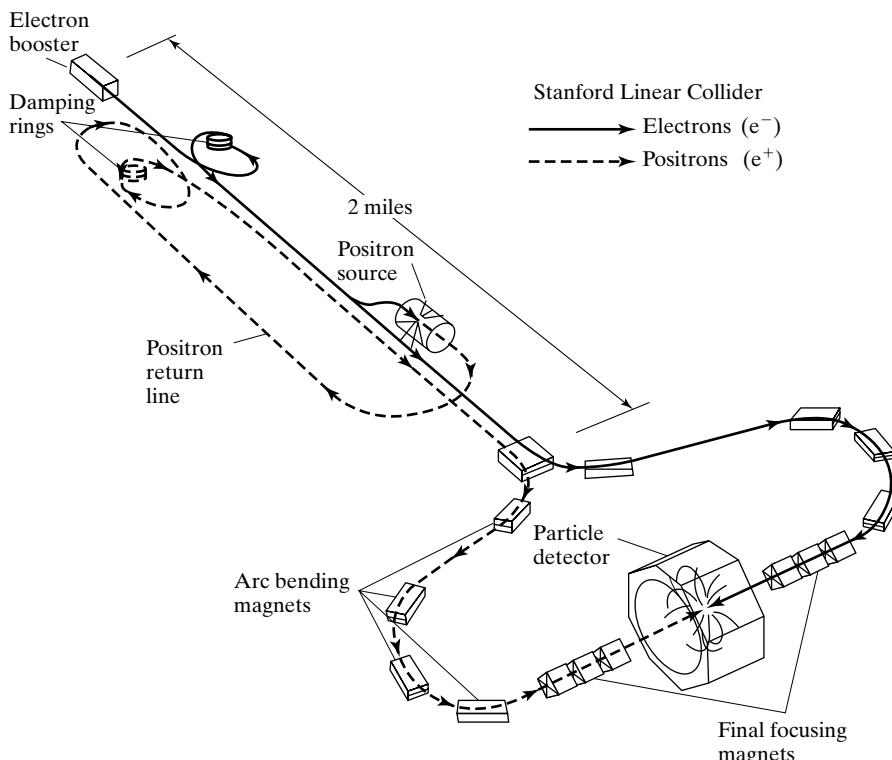
Linacs

We described a low-energy linac in Section 17.11. Most high-energy linacs work in a quite different way: Instead of giving the particle discrete boosts in the gaps in a series of drift tubes, these machines accelerate the particle continuously with the electric field of an electromagnetic wave traveling along the evacuated pipe. The charged particles ride along with the wave, much as a surfer rides a water wave.

To achieve high energies, linacs must be very long. For example, the world’s highest-energy linac, the Stanford Linear Accelerator, which accelerates electrons and positrons to 50 GeV, is 2 miles long. [For comparison, the circular proton synchrotron at the Fermi National Accelerator Laboratory, which accelerates protons to 1000 GeV (or 1 TeV), has a diameter of 1.3 miles.] However, in accelerating electrons and positrons, the linac has a crucial advantage over circular machines: Because of their small mass, electrons and positrons emit intense electromagnetic radiation when forced to follow a circular path (see Problems 11.7 to 11.9). At extremely high energies (many hundreds of GeV), the resulting loss of energy becomes so great that it is impractical to accelerate electrons or positrons in circular machines, and the linac is the only practical alternative.

The Stanford Linear accelerator, shown in Fig. 18.21, can accelerate electrons and positrons simultaneously — the oppositely charged particles riding on opposite sides of a wave crest as it travels along the tube. As they leave the accelerator, the two kinds of particle are separated by a magnetic field, which bends them in opposite directions. The separated beams are steered around two large arcs, shaped like a pair of giant tongs, and then meet in a head-on collision with total energy up to 100 GeV.



Section 18.11 • High-Energy Accelerators **135****FIGURE 18.21**

The Stanford Linear Collider. Electrons and positrons are simultaneously accelerated to 50 GeV in the 2-mile linac. (The positrons are created by diverting some of the electrons about two-thirds of the way along the linac; these diverted electrons collide with a metal plate and create copious positrons, which are then guided back to the start of the linac. The damping rings near the start of the linac are used to focus the electrons and positrons into dense bunches of suitable shape.) At the end of the linac, magnetic fields separate the electrons and positrons and guide them to a head-on collision at the center of the detector.

Synchrotrons

In Section 17.11 we described the cyclotron, which uses a uniform magnetic field to hold charged particles in an approximately circular orbit. This lets the particles pass repeatedly through the same accelerating electrodes and makes the cyclotron a relatively compact machine. As the particles gain energy, they move faster and the radius of their orbit grows in proportion:

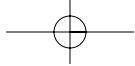
$$R = \frac{p}{qB} = \frac{mv}{qB} \quad [\text{for nonrelativistic motion}] \quad (18.44)$$

Thus the particles spiral steadily outward. However, because R is proportional to v , the period of their orbit ($\tau = 2\pi R/v$) remains constant, and the accelerating fields of a cyclotron can therefore be supplied by AC of fixed frequency.

Unfortunately, the simple proportionality of R and v breaks down once the motion becomes relativistic. Since $p = \gamma mv$, (18.44) must be replaced by

$$R = \frac{p}{qB} = \frac{\gamma mv}{qB} \quad [\text{relativistic}] \quad (18.45)$$

Thus, the orbital period, $\tau = 2\pi R/v$, begins to change once γ becomes appreciably different from 1. If we continued to drive the machine at constant frequency, the particles would get out of step with the accelerating fields until they ceased to gain energy. There are several ways around this difficulty. One is to vary the frequency of the AC supply to keep it synchronized with the changing orbital frequency. (A cyclotron modified in this way is called a



136 Chapter 18 • Elementary Particles

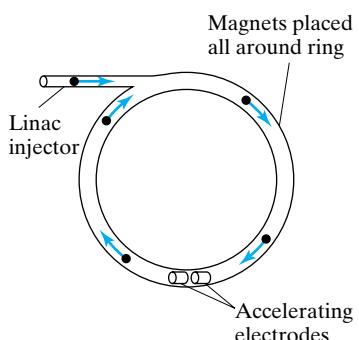


FIGURE 18.22

A synchrotron. Protons (or other charged particles) are injected from a preliminary accelerator, such as a linac, into the ring-shaped evacuated tube. They are held in the circular orbit by magnets placed all around the ring (oriented to make B vertically out of the page in this picture). Each time they go round the ring they are given a boost by the electric field produced by one or more accelerating electrodes. As the particles speed up, the frequency of the accelerating fields must increase to remain synchronized with the orbit, and the magnetic field must increase to hold them in the orbit. The same synchrotron can accelerate protons and antiprotons simultaneously, with the antiprotons circulating in the opposite direction, since the magnetic force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$ is unchanged if we reverse both q and \mathbf{v} .

synchro-cyclotron.) However, a second difficulty develops as we push for higher and higher energies: From (18.45) it follows that $p = qBR$. Thus, the higher the final momentum, the larger must be the outside radius of the circular magnet and the cost of building large enough magnets quickly becomes prohibitive.

In the **synchrotron** the spiral path of the cyclotron is replaced by a circular path of fixed radius, as sketched in Fig. 18.22. The particles move around a ring-shaped, evacuated pipe of radius R . They are repeatedly accelerated by one or more electrodes located around the ring and are held in their circular orbit by magnets placed all around the ring. In this way a magnetic field is required only in the immediate vicinity of the radius R (not for all $r \leq R$), and an enormous saving in cost results. There is, of course, a price for this saving. Since $p = qBR$, and since R is fixed, the B field has to be increased as p increases. In addition, the orbital period, $\tau = 2\pi R/v$, gets shorter as v increases; thus the frequency with which the accelerating fields are applied must also increase as the particles speed up. Thus, in a synchrotron, both B and f must be varied in step with the speed of the particles. One example of a synchrotron is the Super Proton Synchrotron (SPS) at CERN near Geneva, Switzerland, where the W and Z particles were discovered, as described in Section 18.9; another is the Tevatron in Illinois, shown in Fig. 18.23.

Since $p = qBR$, the maximum momentum of the particles is limited by the maximum attainable magnetic field and the radius of the ring: $p_{\max} = qB_{\max}R$. In the last decade or two it has become possible to achieve very large B fields by using superconducting magnets (that is, electromagnets whose current flows in superconducting wires). This was first successfully accomplished with the Tevatron (Fig. 18.23), which was originally built with conventional magnets (B_{\max} around 2 T) and could accelerate protons to around 400 GeV. With the introduction of superconducting magnets (B_{\max} around 4 T), this machine can now reach a little more than 1000 GeV using the same underground tunnel that houses the original ring. (It is because 1000 GeV = 1 TeV that the machine is now called the Tevatron.) A second important advantage of superconducting magnets is that they waste far less energy in Joule heating.

FIGURE 18.23

Aerial view of the Tevatron at the Fermi National Accelerator Laboratory in Illinois, which accelerates protons and antiprotons to about 1 TeV (that is, 1000 GeV). The large circle is the main synchrotron ring and is more than a mile in diameter. The building on the left is 16 stories high.



Example 18.2

If the maximum B field in a synchrotron is 3 T, what must be the machine's circumference if it is to accelerate protons to 1 TeV? How long will the protons take to make one orbit once they reach full energy?

At 1 TeV the rest energy is such a small fraction of the total energy that we can take $E = pc$. Therefore,

$$R = \frac{p}{eB} = \frac{E}{ceB} = \frac{(10^{12} \text{ eV}) \times (1.6 \times 10^{-19} \text{ J/eV})}{(3 \times 10^8 \text{ m/s}) \times (1.6 \times 10^{-19} \text{ C}) \times (3 \text{ T})} = 1.1 \times 10^3 \text{ m}$$

$$\approx 1 \text{ km},$$

and the circumference is about 6 km (which is in fact the circumference of the Tevatron in Fig. 18.23).

At full energy (1 TeV), v is very close to c , and

$$\tau = \frac{2\pi R}{c} = \frac{2\pi \times (1.1 \times 10^3 \text{ m})}{3 \times 10^8 \text{ m/s}} = 2.3 \times 10^{-5} \text{ s}$$

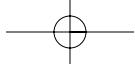
At this rate the protons make some 40,000 trips around the 6-km circumference in each second. In fact, the protons are injected into the synchrotron ring with an energy of 150 GeV, and they are already traveling close to c , so the orbital frequency changes very little during this final acceleration process.

At present there are several very high-energy accelerators under construction or being designed. One of these, the LHC (large hadron collider) at the CERN laboratory in Geneva will produce two beams of 7 TeV each, which will collide head-on to give a total of 14 TeV to produce new particles. A somewhat different development has been the building of machines that accelerate heavy ions (instead of "elementary particles" such as protons or electrons). For example, the Relativistic Heavy Ion Collider (RHIC) at Brookhaven Lab on Long Island, New York, accelerates two beams of bare gold nuclei to an energy of 20 TeV* each, in the hope that their head-on collision will create a "quark-gluon plasma," such as would have existed in the first moments after the big bang. It is the result of one such collision at RHIC that is shown in the picture on the front cover of this book.

CHECKLIST FOR CHAPTER 18

CONCEPT	DETAILS
Antiparticles positron	Same mass and spin as particle, but opposite charge, etc. (Sec. 18.2) antielectron
Force carriers mass-range relation	Photon (for electromagnetic force), pion (for nuclear force) $R = \hbar/2mc$ (18.11)
The four fundamental interactions	Strong, electromagnetic, weak, and gravitational (Sec. 18.4)

*This is the total energy of one gold nucleus and is, in some ways, misleading. In fact it is usual to quote the energy *per nucleon* — about 100 GeV for each of the 197 nucleons in any one nucleus.



138 Chapter 18 • Elementary Particles

Leptons	Spin- $\frac{1}{2}$ particles, immune to the strong force (Sec. 18.5) e, μ , τ and the corresponding neutrinos
Hadrons	Particles subject to the strong force (Sec. 18.6) p, n, π , etc.
	Hadron = any particle made up of 3 quarks (a baryon) or a quark–antiquark pair (meson) (Sec. 18.7)
Resonance particles	Any particle too short-lived to leave a visible track; detected as a peak in a suitable graph (Sec. 18.6)
Conservation of baryon number	Sec. 18.6
Quark model	Quark = spin- $\frac{1}{2}$ constituent of hadrons (Sec. 18.7) fractional charge of quarks multiplets and supermultiplets generations (u, d), (s, c), (b, t)
QCD and gluons	Quantum chromodynamics = theory of strong forces among quarks (Sec. 18.8) color charge and confinement.
Electroweak theory	Unified theory of EM and weak interactions (Sec. 18.9) force carriers = γ , W, and Z
High-energy accelerators*	Linacs and synchrotrons (Sec. 18.11)

PROBLEMS FOR CHAPTER 18

SECTION 18.1 (Elementary Particles: The Story So Far)

- 18.1** • To measure distances of order d with quanta of wavelength λ requires $\lambda \lesssim d$. Find the minimum energies of photons required to measure d with (a) $d \approx 10^{-4}$ m, (b) $d \approx 10^{-8}$ m, and (c) $d \approx 10^{-12}$ m.
- 18.2** • To investigate distances of order 10^{-15} m requires a probe whose wavelength λ is of this same order. (a) Supposing that the probe particle has mass m , derive an expression for its total (relativistic) energy in terms of λ . (b) Show that the minimum energy that gives $\lambda \approx 10^{-15}$ m is of order 1 GeV = 10^9 eV.

- 18.3** • To “see” spatial details of a size d , you need a probe (light, for example) whose wavelength satisfies $\lambda \lesssim d$. What energy of photons do you need to “see” spatial details of (a) atomic size ($d \sim 10^{-10}$ m), (b) nuclear size ($d \sim 10^{-14}$ m), (c) the distribution of “sub-elementary” constituents inside a proton (distances of order 10^{-17} m, say). [Your answer to part (a) may surprise you, since it suggests that energies of order keV’s are needed to probe an atom, whereas we know that we can learn about atomic energy levels with photons of a few eV. However, the latter tell us about *energy levels*, but not about *spatial* details. It remains true that to learn about spatial details one needs keV’s.]

- 18.4** •• (a) Given particles of total relativistic energy E and mass m , what is the minimum distance, d_{\min} ,

that can be probed with these particles? (b) For given E , what mass m would give the best resolution? (c) If $E = 1$ TeV = 10^{12} eV, what is the best possible resolution?

SECTION 18.2 (Antiparticles)

- 18.5** • Consider a positron colliding with a stationary electron. Use conservation of energy and momentum to prove that the e^+ and e^- cannot annihilate into a single photon. [This explains why Eq. (18.2) was written as $e^+ + e^- \rightarrow$ photons, with “photons” in the plural.]
- 18.6** • The first antiprotons to be observed were produced in the reaction



which required a minimum kinetic energy of 6 GeV. Below are listed three reactions that would require less energy, if they were possible. Explain why none of them is possible.

- (a) $p + p \rightarrow p + \bar{p}$
(b) $p + p \rightarrow p + p + \bar{p}$
(c) $p + p \rightarrow p + e^+ + e^- + \bar{p}$

- 18.7** • Suppose that n particles all have the same rest mass m and all move with the same velocity \mathbf{u} . Prove that their total energy and total momentum satisfy the Pythagorean relation $E_{\text{tot}}^2 = (p_{\text{tot}}c)^2 + (Mc^2)^2$, appropriate to a single particle of rest mass $M = nm$. (This result is needed in Problem 18.10.)

- 18.8** •• Prove that a single photon (isolated from any other particles) cannot convert spontaneously into an e^+e^- pair:

$$\gamma \not\rightarrow e^+ + e^-$$

[Hint: Use the Pythagorean relation (2.23) to relate the energy and momentum on each side of this reaction, and then show that the reaction cannot conserve both energy and momentum. This result explains the need for the nucleus in Eq. (18.1); the recoil of the nucleus allows energy and momentum to be conserved.]

- 18.9** •• To identify the antiproton, its discoverers found its mass by separately measuring its momentum and speed. The many particles produced in the original collisions were sent through a magnetic field chosen to select only particles with momentum $1.2 \text{ GeV}/c$ and charge $-e$. These particles were then timed in their flight between two scintillation counters some 12 m apart. (a) How long would an antiproton with $p = 1.2 \text{ GeV}/c$ take to travel this distance? (b) To get an idea of the difficulties of the experiment, calculate how long a π^- ($m = 140 \text{ MeV}/c^2$) would take to travel the same distance. (c) If the experimenters could measure flight times to about a nanosecond, could they tell the difference between a \bar{p} and a π^- in this way?

- 18.10** •• The first antiprotons to be observed were produced in the reaction (18.46) (Problem 18.6), with one of the initial protons at rest. The minimum value of the incident kinetic energy K_i needed to induce the reaction is called the threshold energy for the reaction and is surprisingly large compared to $2mc^2$. To calculate this threshold energy, note that the minimum of K_i occurs when the final kinetic energy is as small as possible, to maximize the fraction of K_i available to create particles. This requires that all final particles have the same velocity, so that there is no kinetic energy “wasted” in relative motion. Write down the Pythagorean relation for the total energy and momentum of the final particles. (Use the result of Problem 18.7.) Then use conservation of energy and momentum to rewrite this relation in terms of the energy and momentum, E_i and p_i , of the incident projectile. Finally, use the Pythagorean relation for E_i and p_i to eliminate p_i . Show that $E_i = 7mc^2$ and hence $K_i = 6mc^2$.

- 18.11** •• In Problem 18.10 the minimum (or threshold) energy E_i for a proton incident on a stationary proton to produce an antiproton in the reaction (18.46) is shown to be $7mc^2$. A more sophisticated and slicker way to derive the same result uses the Lorentz transformation for energy of a particle. This was derived in Problem 2.14, and the relevant part of the result is this: If a particle has energy E and momentum \mathbf{p} in a frame S , then its energy in a second frame S' , (traveling with velocity v along the x axis) is

$$E' = \gamma(E - vp_x)$$

Consider the process (18.46) as seen in a frame S in which the incident protons approach one another along the x axis with equal but opposite momenta. (This frame is called the **center of momentum** —

or CM — frame.) In this CM frame it is obvious that the minimum total incident energy is $4mc^2$, for which all of the final particles are at rest. (a) Write down the energy and momentum of each of the incident protons in this frame. (b) The experiment is usually carried out in the **lab frame** S' , where one of the incident protons is at rest. Using the Lorentz transformation for energy, find the energies of the two incident protons in the lab frame, showing that one is mc^2 and the other $7mc^2$.

SECTION 18.3 (Pions and Muons)

- 18.12** • Equation (18.11), $R \approx \hbar/(2mc)$ relates the range R of a force to the mass m of the particle that carries the force. (a) Apply this relation to the electromagnetic force, whose carrier is the massless photon. (Note that because electromagnetic forces obey an inverse-square law they can, in principle, be felt at any distance, however large.) (b) It is now known that the so-called weak force is carried by a particle called the W and that the range of the weak force is of order 10^{-18} m . Estimate the mass of the W particle.

- 18.13** • The force between two nucleons has a short-range repulsive component that is believed to be carried by a particle called the ω meson, whose mass is $783 \text{ MeV}/c^2$. Estimate the range of this force.

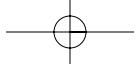
- 18.14** • The force between two nucleons has an attractive component which can be thought of as carried by a σ meson with mass about $550 \text{ MeV}/c^2$. Estimate the range of this force.

- 18.15** • Find the maximum time Δt for which a nucleon can emit a virtual pion, $N \rightarrow N + \pi$.

- 18.16** • Yukawa's theory of the strong nuclear force requires that a nucleon be able to emit a virtual pion, $N \rightarrow N + \pi$. Explain why the pion spin, s_π , must be an integer. [Hint: The time-energy uncertainty relation allows the nonconservation of energy in this virtual process, but the same is not true of angular momentum; that is, total angular momentum is always conserved, even in virtual processes.]

- 18.17** • Explain why one would never observe a process in which a massive particle x emits a particle y , while x itself is completely unchanged, $x \rightarrow x + y$.

- 18.18** • Numerous pions are created by cosmic rays in the upper atmosphere, but very few reach sea level. One reason is that most of the pions decay before they reach sea level, and another is that they are quickly absorbed by the nuclei in the atmosphere. To estimate this second effect, answer the following: (a) About how many nuclei (of nitrogen and oxygen) are contained in a column of atmosphere of area 1 m^2 , extending from sea level to the top of the atmosphere? (Atmospheric pressure at sea level is about 10^5 N/m^2 .) (b) About how many of these nuclei are needed to completely cover 1 m^2 in a layer one nucleus thick? [Recall Eq. (16.3) for the nuclear radius.] (c) Use these two results to estimate the number of nuclei through which a pion would have to pass traveling from the upper atmosphere down to sea level.



140 Chapter 18 • Elementary Particles

18.19 •• Consider a negative pion caught in a 1s orbit around an oxygen nucleus. Because the pion interacts strongly with nucleons, it will be absorbed promptly if there is any probability of its penetrating the nucleus. Show that the probability of finding the pion inside the nucleus is about 0.25% (which is ample to ensure that the pion is rapidly absorbed). [Hint: The radial probability density is given by Eq. (8.85), with the constant A given by Eq. (8.88), except that $a_B = \hbar^2/(m_e k e^2)$ must be replaced by $a = \hbar^2/(m_\pi Z k e^2)$. The required probability is that of finding the pion in the range $0 \leq r \leq R \approx 3$ fm. Notice that $R \ll a$, which lets you approximate the factor $e^{-2r/a}$ by 1 inside the integral.]

SECTION 18.4 (The Four Fundamental Forces)

18.20 • In Section 18.4 it is stated that the gravitational force between two nuclear particles is negligible compared with the electrostatic force. Verify this claim by computing the ratio $F_{\text{grav}}/F_{\text{elec}}$ for the forces between two stationary protons.

18.21 • To contrast the strengths of the strong and electromagnetic interactions between two protons, do the following: Find the nuclear potential energy of two protons 1 fm apart (see Fig. 16.3). Calculate the electrostatic potential energy of the same two protons and compare.

18.22 • Of the four fundamental interactions, which is the strongest that can cause each of the following processes?

- (a) $^{12}\text{N} \rightarrow ^{12}\text{C} + e^+ + \nu$
- (b) $\pi^+ + p \rightarrow \pi^+ + p$ (elastic scattering)
- (c) $e + p \rightarrow e + p$
- (d) $\nu + e \rightarrow \nu + e$
- (e) $e^- + e^+ \rightarrow e^- + e^+$
- (f) $e^- + e^+ \rightarrow \text{photons}$
- (g) $n \rightarrow p + e^- + \bar{\nu}$
- (h) $p + n \rightarrow p + n + \pi^0$
- (i) (excited atom) \rightarrow (ground state) + γ

18.23 • Of the four fundamental interactions, which is the strongest that can cause each of the following processes?

- (a) atom + $\gamma \rightarrow$ (excited atom)
- (b) $p + p + \pi^- \rightarrow p + n$
- (c) $^{14}\text{C} \rightarrow ^{14}\text{N} + e^- + \bar{\nu}$
- (d) $e^- + e^- \rightarrow e^- + e^-$ (elastic scattering)
- (e) $\nu + p \rightarrow \nu + p$
- (f) $\pi^+ \rightarrow \mu^+ + \nu$
- (g) $\pi^- + p \rightarrow \pi^- + p$

SECTION 18.5 (Leptons)

18.24 • The charged pion π^- usually decays into a muon and a neutrino,

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$$

but occasionally into an electron and a neutrino,

$$\pi^- \rightarrow e^- + \bar{\nu}_e$$

The relative frequency of the electron decay (compared to the muon decay) is on the order of 1 in 10^4 . The large difference between the probabilities for these two decays can be explained in terms of the $\mu - e$ mass difference as follows: The theory of weak interactions predicts that the probability for either decay is proportional to $(1 - v/c)$, where v is the speed of the outgoing μ^- or e^- . Calculate the quantity $(1 - v/c)$ for each decay, and compute their ratio. Note that this ratio has the same order of magnitude as the observed relative frequency. (Use the result of Problem 2.40; $m_\pi \approx 140$, $m_\mu \approx 106$, and $m_e \approx 0.51 \text{ MeV}/c^2$.)

18.25 • The electron is the lightest known charged particle. Explain why this means that an isolated electron is perfectly stable.

18.26 •• It appears that there are three conserved quantities, N_e , N_μ , and N_τ , called **lepton numbers**. These are analogous to nucleon number and are defined as follows: The e^- and $\bar{\nu}_e$ have $N_e = 1$ (each), the e^+ and $\bar{\nu}_e$ have $N_e = -1$, and all other particles have $N_e = 0$. If we define N_μ and N_τ correspondingly, then as far as is known, the total value of each of these quantities is always conserved. (a) Verify this claim for the processes $n \rightarrow p + e^- + \bar{\nu}_e$, $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$ and $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$. (b) Use the laws to fill in the blanks in the following decays (where each blank is some kind of neutrino or antineutrino):

$$\begin{aligned} \tau^+ &\rightarrow e^+ + \underline{\quad} + \underline{\quad} \\ \tau^- &\rightarrow \mu^- + \underline{\quad} + \underline{\quad} \end{aligned}$$

SECTION 18.6 (Discovery of More Hadrons)

18.27 • A Σ^+ particle is formed in a bubble chamber in a magnetic field of 0.4 T and leaves a track whose radius of curvature is measured to be $R = 0.5$ m. The charge of the Σ^+ is known to be $+e$. What is its momentum, in $\text{kg} \cdot \text{m/s}$ and in MeV/c ? [Remember the relation (2.48) between R and the momentum.]

18.28 • (a) The measured width in the rest energy of the $\psi(3097)$ meson is $\Gamma = 0.068 \text{ MeV}$. What is this meson's mean lifetime? **(b)** What is the mean life of the $\Sigma(1385)$ baryon, for which $\Gamma = 37 \text{ MeV}$?

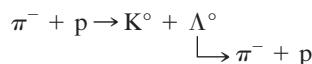
18.29 • Use the time-energy uncertainty relation in the form (18.35) to find the uncertainty in the rest energies (full width Γ) of the four particles listed in Table 18.6.

18.30 • The neutron is unstable with a mean life of about 15 min. (This is the mean life, not the half life.) What is the intrinsic uncertainty Γ in its rest energy? Comment.

18.31 • The Σ^+ baryon has a mean life $\tau = 8.0 \times 10^{-11} \text{ s}$. What is the full width Γ of the intrinsic uncertainty in its rest energy? The present experimental uncertainty in the Σ^+ mass is $0.06 \text{ MeV}/c^2$; does the intrinsic uncertainty affect the measurement of the Σ^+ mass significantly?

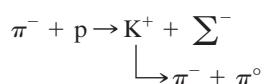
- 18.32** • The Λ baryon has mean life $\tau = 2.6 \times 10^{-10}$ s. What is the intrinsic uncertainty (full width Γ) in its rest energy? The *experimental* uncertainty in the mass is 0.006 MeV/ c^2 ; does the intrinsic uncertainty affect the measurement of the Λ mass significantly?

- 18.33** • Consider the reaction

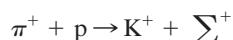


We know that the proton has baryon number $B = +1$ and that pions have $B = 0$. Given that baryon number is always conserved, deduce the baryon number of the Λ° and the K° .

- 18.34** •• (a) Consider the reaction



Knowing that the proton is a baryon and the pions are not, deduce the baryon numbers B of the K^+ and Σ^- . (b) Given the reaction



deduce the baryon number of the Σ^+ . (c) Is the Σ^- the antiparticle of the Σ^+ ?

- 18.35** ••• Figure 18.24 shows the track of a K^+ meson in a bubble chamber. The K^+ has decayed into a μ^+ and an invisible neutrino: $K^+ \rightarrow \mu^+ + \nu_\mu$. The paths of the charged particles were curved by a magnetic field $B = 0.50$ T directed out of the page. The measured radii of the curved paths are $R_K = 1.0$ m for the K^+ and $R_\mu = 1.7$ m for the μ^+ . (a) Use this information to find the momenta of the K^+ and μ^+ . [Recall Eq. (2.48), relating R to p .] (b) The angle between \mathbf{p}_K and \mathbf{p}_μ at the moment of decay is 71° , as shown. Deduce the momentum of the invisible neutrino. (c) Assuming the known masses of the μ^+ and ν_μ ($m_\mu = 106$ MeV/ c^2 and $m_\nu \approx 0$), calculate their energies. (d) Deduce the energy of the K^+ , and hence find its mass and speed.

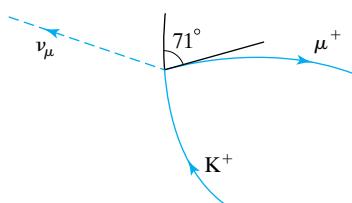


FIGURE 18.24

(Problem 18.35).

SECTION 18.7 (The Quark Model of Hadrons)

- 18.36** • According to the quark model, the reaction $\pi^+ + p \rightarrow K^+ + \Sigma^+$ should be understood as follows:

$$(u\bar{d}) + (uud) \rightarrow (u\bar{s}) + (uus)$$

Rewrite each of the following reactions in the same way

- (a) $\pi^+ + n \rightarrow K^\circ + \Sigma^+$
- (b) $K^+ + n \rightarrow K^\circ + p$
- (c) $\Delta^{++} \rightarrow p + \pi^+$
- (d) $e^- + p \rightarrow e^- + \Delta^{++} + \pi^-$

(The quark constituents of all the hadrons concerned can be found in Tables 18.5 to 18.7.)

- 18.37** • Analyze the following reactions in terms of the quark constituents of the hadrons involved, as in Problem 18.36:

- (a) $\pi^- + p \rightarrow K^+ + \Sigma^-$
- (b) $\Sigma^- \rightarrow n + \pi^-$
- (c) $e^- + n \rightarrow e^- + \Delta^\circ$
- (d) $\Delta^\circ \rightarrow \pi^- + p$

- 18.38** • The Ξ° (mass 1315 MeV/ c^2) is a “doubly strange” baryon with composition (uss) and is a member of a doublet of two baryons with almost exactly the same mass. Based on this information, what are the quark composition of the other member and its charge, and what should be its symbol?

- 18.39** • The Σ^+ is a strange baryon containing one s quark, $\Sigma^+ = (uus)$ and is one of a triplet of particles with different charges but almost exactly the same masses. Given this information, what would you suggest for the quark constituents of the two companions of the Σ^+ ? What should they be called, and what are their charges?

- 18.40** • Use Fig. 18.13 to find the energy of the photon emitted in each of the decays of the excited states of the $(c\bar{c})$ meson. What sort of photons are these? (Visible? UV? etc.)

SECTION 18.8 (The Strong Force and QCD)

- 18.41** • Figure 18.25 shows a proton and Σ^+ interacting by exchange of a K° meson. Redraw this picture in detail [as in Fig. 18.17(b)] to show the quark constituents of all the particles involved.

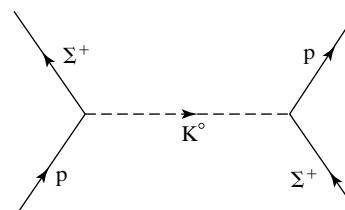


FIGURE 18.25

(Problems 18.41 and 18.42).

- 18.42** • Figure 18.25 shows a proton and Σ^+ interacting by exchange of a K° meson. (a) Draw the corresponding diagram in which a proton and Σ° exchange a K^+ . (b) Redraw your diagram to show the quark constituents of all the particles involved.

142 Chapter 18 • Elementary Particles

SECTION 18.9 (Electroweak Interactions: The W and Z Particles)

18.43 Given that the weak force is carried by the W particle with mass about $80 \text{ GeV}/c^2$, estimate the range of the weak force.

18.44 In a fixed-target experiment the target is usually a liquid or solid with a high density of individual target particles. In a colliding-beam experiment the “target” is one of the two beams, and the corresponding density is much lower. This means that colliding-beam experiments yield a much smaller number of events and are hence much harder to perform. Estimate the magnitude of this difference as follows: (a) Find the density of protons (number per unit volume) in a liquid hydrogen target. (The mass density of liquid hydrogen $\approx 0.07 \text{ g/cm}^3$.) (b) Estimate the density of protons in a beam with the following characteristics: beam diameter $\approx 1 \text{ mm}$, beam current $\approx 1 \text{ A}$, speed of particles $\approx c$. (c) Compare the “target” densities in the two types of experiment. (The rate of events depends on several other factors as well; nevertheless, low target density is one of the main problems to be overcome in colliding-beam experiments.)

18.45 When a neutron (udd) undergoes β decay, what actually happens is that one of the d quarks emits a virtual W^- , which decays into an e^- and $\bar{\nu}_e$. Draw a Feynman diagram to represent the β decay of a neutron, showing all quark constituents.

18.46 The π^0 decays electromagnetically with $\tau(\pi^0) = 8.7 \times 10^{-17} \text{ s}$, but the π^+ decays only weakly with $\tau(\pi^+) = 2.6 \times 10^{-8} \text{ s}$. You can explain the large difference between these lifetimes in terms of the different ranges of the electromagnetic and weak interactions: First, recall that any meson is composed of two quarks. Next, note that, according to the electroweak theory, the intrinsic strengths of the electromagnetic and weak interactions are the same; this means that the probabilities for either kind of interaction are about the same, *provided* that two quarks are within the relevant range. Now the two quarks in a meson are always within the (infinite) range of the electromagnetic force; on the other hand, they are very seldom within the range of the weak force. Given that the two quarks move more or less randomly inside the volume of the meson (radius of order 1 fm) and that the range of the weak force is of order 10^{-3} fm , estimate what fraction of the time the two quarks are within the range required for weak interaction. Use these considerations to estimate the ratio $\tau(\pi^0)/\tau(\pi^+)$, and show that your answer is of the same order as the observed ratio.

18.47 Lifetimes of mesons that decay via the strong interaction are of order 10^{-23} s . Lifetimes for weak decays vary considerably, depending on the energy released, but a representative value is 10^{-10} s . Explain this difference using an argument parallel to that sketched in Problem 18.46. [Hint: In this problem you are comparing the weak and *strong* interactions. The difference in the intrinsic strengths of these interac-

tions would, by itself, make the strong decays some 10^4 times faster. The two quarks inside the meson are always within the range of the strong color force.]

18.48 Suppose that a particle z is expected to be produced in a reaction $x + y \rightarrow z$. (For example, the Z^0 particle could be produced in the reaction $p + \bar{p} \rightarrow Z^0$.) (a) By far, the simplest experimental arrangement is to fire the x particle at a stationary y target (or vice versa). Prove that in this case the required incident total energy ($E = E_x + m_y c^2$) is equal to

$$E = \frac{m_z^2 - m_x^2 + m_y^2}{2m_y} \quad (\text{y stationary})$$

[Hint: Use conservation of energy and momentum, and apply the Pythagorean relation (2.23) to particles x and z .] (b) A much more efficient arrangement (although much harder experimentally) is to use colliding beams, in which the x and y collide head-on with equal but opposite momenta. Prove that in this case the required total energy ($E = E_x + E_y$) is

$$E = m_z c^2 \quad (\text{colliding beams})$$

(c) Suppose that $m_x = m_y$ (as is the case in pp or $p\bar{p}$ collisions). Show that the particle z that can be produced has

$$m_z c^2 = \sqrt{2m_y c^2 E} \quad (\text{stationary target})$$

for a stationary target, but

$$m_z c^2 = E \quad (\text{colliding beams})$$

for colliding beams. If $E = 400 \text{ GeV}$ and m_x and m_y equal the proton mass, what is the mass m_z that can be produced in each case? (d) If $m_z = 90 \text{ GeV}/c^2$ (about the Z^0 mass) and m_x and m_y equal the proton mass, what is the energy, E , needed to produce the z particle in each case?

SECTION 18.10 (Summary and Outlook)

18.49 The proton is the lightest known baryon. Explain why if we ever observe the decay of a proton (as predicted by certain grand unified theories), baryon number cannot be conserved.

18.50 Certain grand unified theories predict that the proton should be unstable with a lifetime of order 10^{31} years . Estimate how long one would have to wait for one of the protons in a gallon (about 4 liters) of water to decay.

18.51 Experiments are under way to detect the supposed decay of the proton, by monitoring an underground cave full of water. Assuming that the proton's mean life is about 10^{31} years , estimate how much water is needed to give one decay per hour.

18.52 The mean kinetic energy of particles at temperature T is $\langle K \rangle = 3k_B T/2$, where k_B is Boltzmann's constant, $k_B = 8.62 \times 10^{-5} \text{ eV/K}$. (a) Given that mean energies of order 10 keV are sufficient to strip

the electrons off most atoms, find the temperature T_{atom} above which one would find few atoms. (For its first million years or so, the universe was above this temperature and so contained just electrons and nuclei.) **(b)** Given that mean energies of order 1 MeV will knock most nuclei apart, find the temperature T_{nuc} , above which one would find few nuclei. (For its first 3 minutes or so, the universe was above T_{nuc} and contained just electrons, protons, and neutrons.)

SECTION 18.11 (High-Energy Accelerators)

18.53 • In the Tevatron, protons are injected from a ring that uses conventional magnets into the final superconducting ring with about 150 GeV. They then gain roughly 1 MeV per orbit as they move around the final ring. **(a)** Roughly how many orbits does a proton have to make in the final ring to gain the full energy of 1 TeV? **(b)** Given that the ring has radius 1 km, how far does the proton travel inside the final ring, and how long does this final acceleration take?

18.54 • One might hope that the next generation of accelerators after the Large Hadron Collider being constructed at the CERN lab in Switzerland will produce particles with energies of at least 200 TeV (some 30 times more than the LHC). Assuming that the machine is a proton

synchrotron and that the maximum practical B field is 10 T (the same as for the LHC), what would be the required diameter of such a machine? (Your answer should suggest that some completely new technology will probably be needed before a machine of this energy can be undertaken.)

18.55 • The Large Hadron Collider, under construction at the CERN laboratory, will accelerate two beams of protons to 7 TeV each. Given that the beam tunnel is 27 km in circumference, what is the minimum B field needed to hold the protons in their circular orbit?

18.56 • In the proposed superconducting supercollider, which was cancelled by Congress in 1993, protons would have accelerated to 20 TeV, held in an approximately circular orbit by a magnetic field of roughly 5 T. **(a)** Estimate the required diameter of the ring. **(b)** How long would a proton have taken to go once around the ring (at full energy)?

18.57 • We have said that as the particles in a synchrotron speed up, the frequency f of the accelerating field must be increased. Given that protons are injected into the Tevatron ring with 150 GeV and are then accelerated to 1 TeV, what is the percent variation in f required in this acceleration?