CHAPTER 5 - Gravitation

5.1 Introduction

By 1666, Newton had formulated and numerically checked the gravitation law he eventually published in his book Principia in 1687. Newton waited almost 20 years to publish his results because he could not justify his method of numerical calculation in which he considered Earth and the Moon as point masses. With mathematics formulated on calculus (which Newton later invented), we have a much easier time proving the problem Newton found so difficult in the seventeenth century.

Newton's law of universal gravitation states that *each mass particle attracts every other particle in the universe with a force that varies directly as the product of the two masses and inversely as the square of the distance between them*. In mathematical form, we write the law as

$$\mathbf{F} = -G\frac{mM}{r^2}\mathbf{e}_r \tag{5.1}$$

where at a distance r from a particle of mass M a second particle of mass m experiences an attractive force (see Figure 5-1).

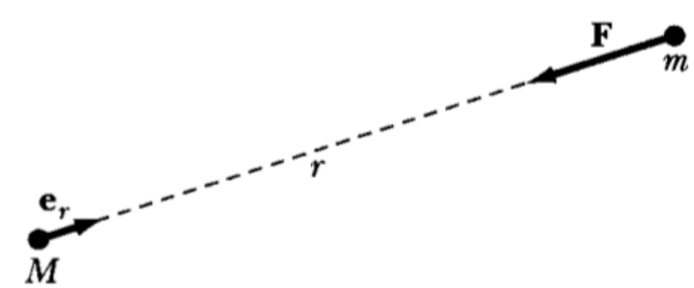


FIGURE 5-1 Particle *m* feels an attractive gravitational force toward *M*.

The unit vector $\mathbf{e}_r$ points from M to m, and the minus sign ensures that the force is attractive - that is, that m is attracted toward M.

A laboratory verification of the law and a determination of the value of G was made in 1798 by the English physicist Henry Cavendish. Cavendish's experiment, described in many elementary physics texts, used a torsion balance with two small spheres fixed at the ends of a light rod. The two spheres were attracted to two other large spheres that could be placed on either side of the smaller spheres.
The official value for G is $6.673 \pm 0.010 \times 10^{-11}$ N m^2/kg^2. Interestingly, although G is perhaps the oldest known of the fundamental constants, we know it with less precision than we know most of the modern fundamental constants such as e, c, and $\hbar$. Considerable research is ongoing today to improve the precision of G.

In the form of Equation 5.1, the law strictly applies only to *point particles*. If one or both of the particles is replaced by a body with a certain extension, we must make an additional hypothesis before we can calculate the force. We must assume that the gravitational force field is a *linear field*. In other words, we assume that it is possible to calculate the net gravitational force on a particle due to many other particles by simply taking the vector sum of all the individual forces. For a body consisting of a continuous distribution of matter, the sum becomes an integral (Figure 5-2):

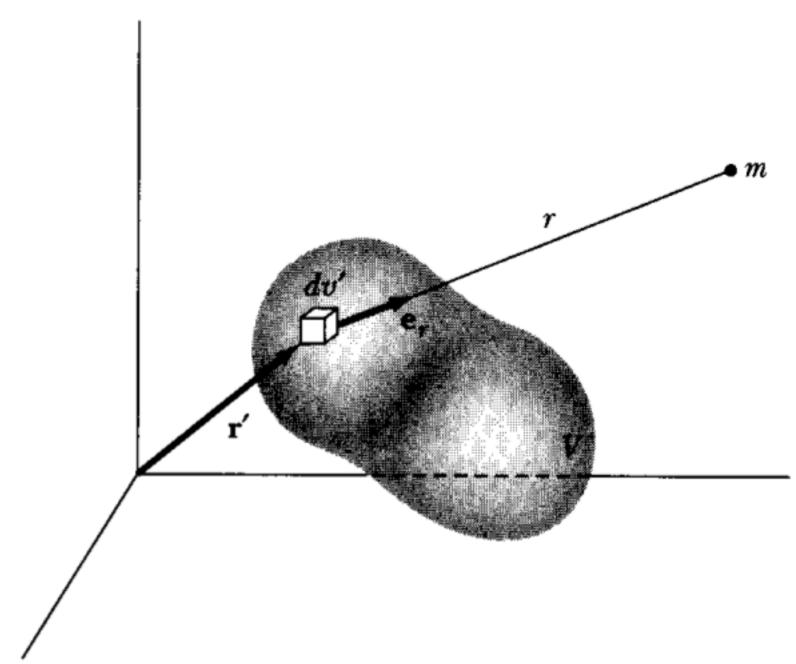


FIGURE 5-2 To find the gravitational force between a point mass m and a continuous distribution of matter, we integrate the mass density over the volume.

$$\mathbf{F} = -Gm\int_V \frac{\rho(\mathbf{r}')\mathbf{e}_r}{r^2}dv' \tag{5.2}$$

where $\rho(\mathbf{r}')$ is the mass density and dv' is the element of volume at the position defined by the vector **r**' from the (arbitrary) origin to the point within the mass distribution.

If both the body of mass M and the body of mass m have finite extension, a second integration over the volume of m will be necessary to compute the total gravitational force.

The **gravitational field vector g** is the vector representing the force per unit mass exerted on a particle in the field of a body of mass M. Thus

$$\mathbf{g} = \frac{\mathbf{F}}{m} = -G\frac{M}{r^2}\mathbf{e}_r \tag{5.3}$$

or

$$\boxed{\mathbf{g} = -G\int_V \frac{\rho(\mathbf{r}')\mathbf{e}_r}{r^2}dv'} \tag{5.4}$$

Note that the direction of $\mathbf{e}_r$ varies with r' (in Figure 5-2).

The quantity **g** has the dimensions of force per unit mass, also equal to acceleration. In fact, near the surface of the earth, the magnitude of **g** is just the quantity that we call the gravitational acceleration constant. Measurement with a simple pendulum (or some more sophisticated variation) is sufficient to show that |**g**| is approximately 9.80 m/s^2 (or 9.80 N/kg) at the surface of the earth.

5.2 Gravitational Potential

The gravitational field vector **g** varies as $1/r^2$ and therefore satisfies the requirement ($\nabla \times \mathbf{g} = 0$) that permits **g** to be represented as the gradient of a scalar function. Hence, we can write

$$\boxed{\mathbf{g} \equiv -\nabla\Phi} \tag{5.5}$$

where Φ is called the gravitational potential and has dimensions of (*force per unit mass*) x (*distance*), or *energy per unit mass*.

Because **g** has only a radial variation, the potential Φ can have at most a variation with r. Therefore, using Equation 5.3 for **g**, we have

$$\nabla\Phi = \frac{d\Phi}{dr}\mathbf{e}_r = G\frac{M}{r^2}\mathbf{e}_r$$

Integrating, we obtain

$$\boxed{\Phi = -G\frac{M}{r}} \tag{5.6}$$

The possible constant of integration has been suppressed, because the potential is undetermined to within an additive constant; that is, only differences in potential are meaningful, not particular values. We usually remove the ambiguity in the value of the potential by arbitrarily requiring that $\Phi \to 0$ as $r \to \infty$; then Equation 5.6 correctly gives the potential for this condition.

The potential due to a continuous distribution of matter is

$$\Phi = -G\int_V \frac{\rho(\mathbf{r}')}{r}\,dv' \tag{5.7}$$

Similarly, if the mass is distributed onlv over a thin shell (i.e., a *surface* distribution), then

$$\Phi = -G\int_S \frac{\rho_s}{r}\,da' \tag{5.8}$$

where ρ is the surface density of mass (or *areal mass density*).

Finally, if there is a line source with linear mass density ρ_l , then

$$\Phi = -G\int_\Gamma \frac{\rho_l}{r}\,ds'$$

The physical significance of the gravitational potential function becomes clear if we consider the work per unit mass dW' that must be done by an outside agent on a body in a gravitational field to displace the body a distance $d\mathbf{r}$. In this case, work is equal to the scalar product of the force and the displacement. Thus, for the work done on the body per unit mass, we have

$$\begin{aligned} dW' &= -\mathbf{g}\cdot d\mathbf{r} = (\nabla\Phi)\cdot d\mathbf{r} \\ &= \sum_i \frac{\partial\Phi}{\partial x_i}\,dx_i = d\Phi \end{aligned} \tag{5.10}$$

because Φ is a function only of the coordinates of the point at which it is measured:

$$\Phi = \Phi(x_1, x_2, x_3) = \Phi(x_i)$$

Therefore the amount of work per unit mass that must be done on a body to move it from one position to another in a gravitational field is equal to the difference in potential at the two points.

If the final position is farther from the source of mass M than the initial position, work has been done *on* the unit mass. The positions of the two points are arbitrary, and we may take one of them to be at infinity.

If we define the potential to be zero at infinity, we may interpret Φ at any point to be the work per unit mass required to bring the body from infinity to that point. The *potential energy* is equal to the mass of the body multiplied by the potential Φ. If U is the potential energy, then

$$U = m\Phi \tag{5.11}$$

and the force on a body is given by the negative of the gradient of the potential energy of that body,

$$\mathbf{F} = -\nabla U \tag{5.12}$$

which is just the expression we have previously used (Equation 2.88).

We note that both the potential and the potential energy *increase* when work is done *on* the body. (The potential, according to our definition, is always negative and only approaches its maximum value, that is, zero, as r tends to infinity.)

A certain potential energy exists whenever a body is placed in the gravitational field of a source mass. This potential energy resides in the *field* but it is customary under these circumstances to speak of the potential energy "of the body". We shall continue this practice here. We may also consider the source mass itself to have an intrinsic potential energy. This potential energy is equal to the gravitational energy released when the body was formed or, conversely, is equal to the energy that must be supplied (i.e., the work that must be done) to disperse the mass over the sphere at infinity. For example, when interstellar gas condenses to form a star, the gravitational energy released goes largely into the initial heating of the star As the temperature increases, energy is radiated away as electromagnetic radiation. In all the problems we treat, the structure of the bodies is considered to remain unchanged during the process we are studying. Thus, there is no change in the intrinsic potential energy, and it may be neglected for the purposes of whatever calculation we are making.

EXAMPLE 5.1

What is the gravitational potential both inside and outside a spherical shell of inner radius b and outer radius a?

Solution. One of the important problems of gravitational theory concerns the calculation of the gravitational force due to a homogeneous sphere. This problem is a special case of the more general calculation for a homogeneous spherical shell. A solution to the problem of the shell can be obtained by directly computing the force on an arbitrary object of unit mass brought into the field, but it is easier to use the potential method.

We consider the shell shown in Figure 5-3 and calculate the potential at point P a distance R from the center of the shell.

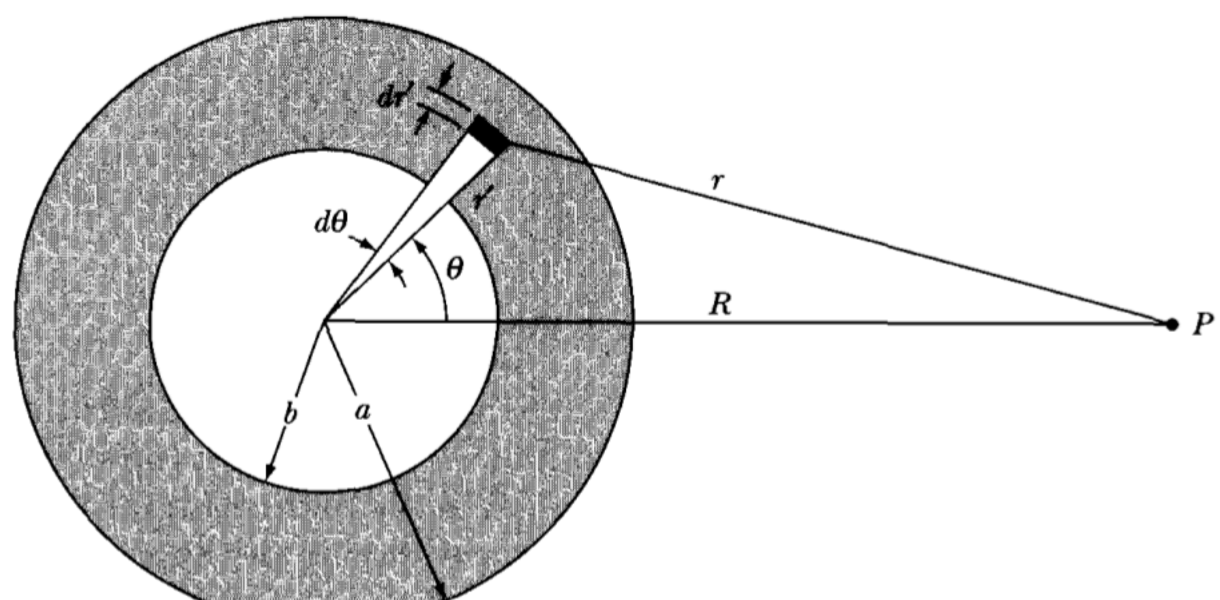


FIGURE 5-3 The geometry for finding the gravitational potential at point *P* due to a spherical shell of mass.

Because the problem has symmetry about the line connecting the center of the sphere and the field point P, the azimuthal angle ϕ is not shown in Figure 5-3 and we can immediately integrate over $d\phi$ in the expression for the potential. Thus,

$$\begin{aligned}\Phi &= -G\int_V \frac{\rho(r')}{r} dv' \\ &= -2\pi\rho G\int_b^a r'^2 dr' \int_0^\pi \frac{\sin\theta}{r} d\theta\end{aligned} \tag{5.13}$$

where we have assumed a homogeneous mass distribution for the shell, $\rho(r') = \rho$. According to the law of cosines,

$$r^2 = r'^2 + R^2 - 2r'R\cos\theta \tag{5.14}$$

Because R is a constant, for a given r' we may differentiate this equation and obtain

$$2r\,dr = 2r'R\sin\theta\,d\theta$$

or

$$\frac{\sin\theta}{r} d\theta = \frac{dr}{r'R} \tag{5.15}$$

Substituting this expression into Equation 5.13, we have

$$\Phi = -\frac{2\pi\rho G}{R}\int_b^a r'\,dr' \int_{r_{\text{min}}}^{r_{\text{max}}} dr \tag{5.16}$$

The limits on the integral over dr depend on the location of point P. If P is *outside* the shell, then

$$\begin{aligned}\Phi(R > a) &= -\frac{2\pi\rho G}{R}\int_b^a r'\,dr' \int_{R-r'}^{R+r'} dr \\ &= -\frac{4\pi\rho G}{R}\int_b^a r'^2\,dr' \\ &= -\frac{4}{3}\frac{\pi\rho G}{R}(a^3 - b^3)\end{aligned} \tag{5.17}$$

But the mass M of the shell is

$$M = \frac{4}{3}\pi\rho(a^3 - b^3) \tag{5.18}$$

so the potential is

$$\boxed{\Phi(R > a) = -\frac{GM}{R}} \tag{5.19}$$

If the field point lies inside the shell, then

$$\begin{aligned}\Phi(R < b) &= -\frac{2\pi\rho G}{R}\int_b^a r'dr'\int_{r'-R}^{r'+R} dr \\ &= -4\pi\rho G\int_b^a r'dr' \\ &= -2\pi\rho G(a^2 - b^2)\end{aligned} \tag{5.20}$$

The potential is therefore constant and independent of position inside the shell.

Finally, if we wish to calculate the potential for points within the shell, we need only replace the lower limit of integration in the expression for $\Phi(R < b)$ by the variable R replace the upper limit of integration in the expression for $\Phi(R > a)$ by R and add the results. We find

$$\begin{aligned}\Phi(b < R < a) &= -\frac{4\pi\rho G}{3R}(R^3 - b^3) - 2\pi\rho G(a^2 - R^2) \\ &= -4\pi\rho G\left(\frac{a^2}{2} - \frac{b^3}{3R} - \frac{R^2}{6}\right)\end{aligned} \tag{5.21}$$

We see that if $R \to a$, then Equation 5.21 yields the same result as Equation 5.19 for the same limit. Similarly, Equations 5.21 and 5.20 produce the same result for the limit $R \to b$. The potential is therefore *continuous*.

If the potential were not continuous at some point, the gradient of the potential - and hence, the force - would be infinite at that point. Because infinite forces do not represent physical reality, we conclude that realistic potential functions must always be continuous.

Note that we treated the mass shell as homogeneous. In order to perform calculations for a solid, massive body like a planet that has a spherically symmetric mass distribution, we could add up a number of shells or, if we choose, we could allow the density to change as a function of radius.

The results of Example 5.1 are very important. Equation 5.19 states that the potential at any point outside of a spherically symmetric distribution of matter (shell or solid, because solids are composed of many shells) is independent of the size of the distribution. Therefore, to calculate the external potential (or the force), we consider all the mass to be concentrated at the center. Equation 5.20 indicates that the potential is constant (and the force zero) anywhere inside a spherically symmetric mass shell. And finally, at points within the mass shell, the potential given by Equation 5.21 is consistent with both of the previous results.

The magnitude of the field vector **g** may be computed from $\mathbf{g} = -d\Phi/dR$ for each of the three regions. The results are

$$\left.\begin{aligned} g(R < b) &= 0 \\ g(b < R < a) &= \frac{4\pi\rho G}{3}\left(\frac{b^3}{R^2} - R\right) \\ g(R > a) &= -\frac{GM}{R^2} \end{aligned}\right\} \qquad (5.22)$$

We see that not only the potential but also the field vector (and hence, the force) are continuous. The derivative of the field vector, however, is not continuous across the outer and inner surfaces of the shell.

All these results for the potential and the field vector can be summarized as in Figure 5-4.

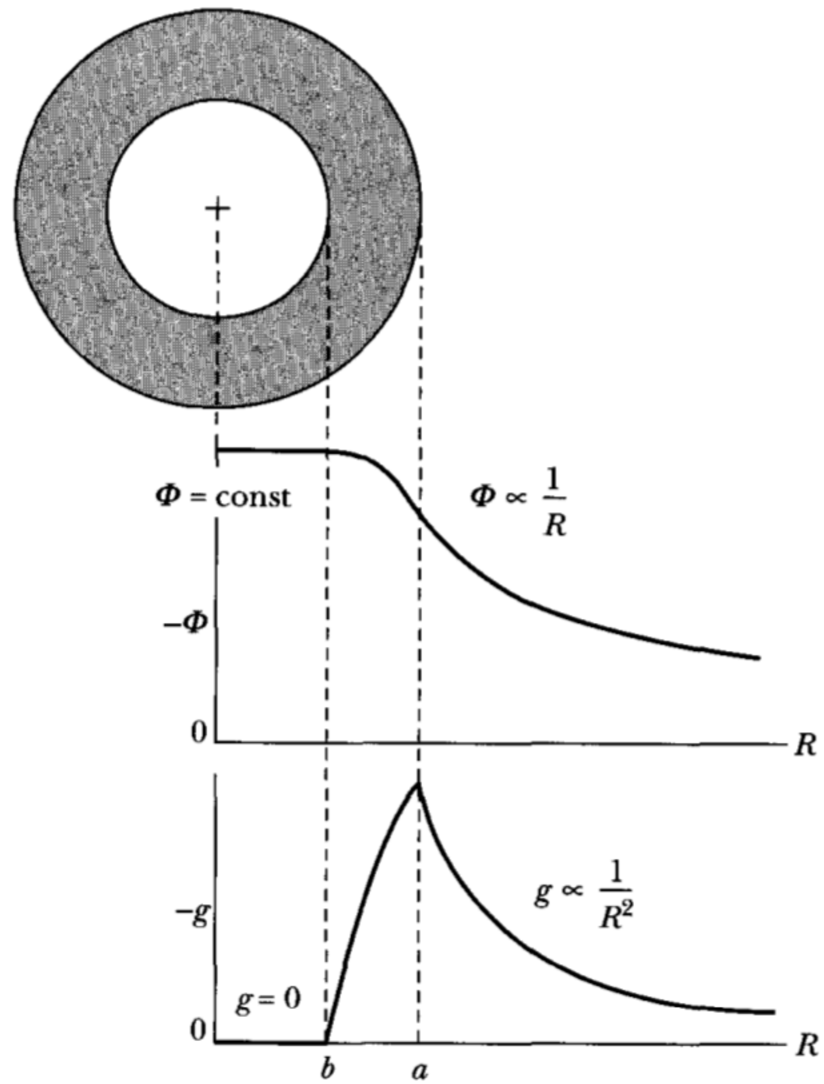


FIGURE 5-4 The results of Example 5.1 indicating the gravitational potential and magnitude of the field vector **g** (actually −**g**) as a function of radial distance.

EXAMPLE 5.2

Astronomical measurements indicate that the orbital speed of masses in many spiral galaxies rotating about their centers is approximately constant as a function of distance from the center of the galaxy (like our own Milky Way and our nearest neighbor Andromeda) as shown in Figure 5-5. Show that this experimental result is inconsistent with the galaxy having its mass concentrated near the center of the galaxy and can be explained if the mass of the galaxy increases with distance R.

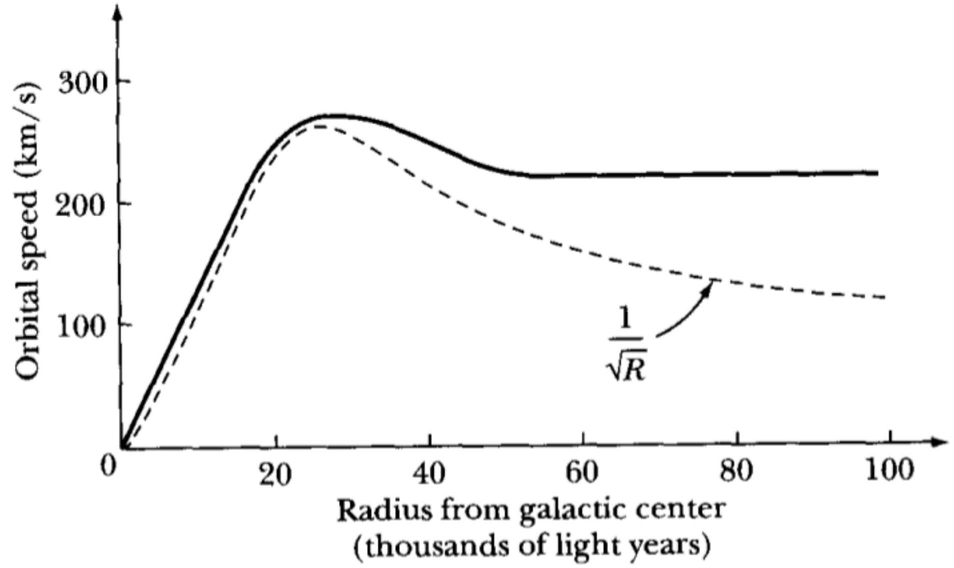


FIGURE 5-5 Example 5.2. The solid line represents data for the orbital speed of mass as a function of distance from the center of the Andromeda galaxy. The dashed line represents the $1/\sqrt{R}$ behavior expected from the Keplerian result of Newton's laws.

Solution. We can find the expected orbital speed v due to the galaxy mass M that is within the radius R. In this case, however, the distance R may be hundreds of light years. We only assume the mass distribution is spherically symmetric. The gravitational force in this case is equal to the centripetal force due to the mass m having orbital speed v.

$$\frac{GMm}{r^2} = \frac{mv^2}{R}$$

We solve this equation for v.

$$v = \sqrt{\frac{GM}{R}}$$

If this were the case, we would expect the orbital speed to decrease as $1/\sqrt{R}$ as shown by the dashed line in Figure 5-5, whereas what is found experimentally is that v is constant as a function of R. This can only happen in the previous equation if the mass M of the galaxy itself is a linear function of R, $M(R) \propto R$. Astrophysicists conclude from this result that for many galaxies there must be matter other than that observed, and that this unobserved matter, often called "dark matter," must account for more than 90 percent of the known mass in the universe. This area of research is at the forefront of astrophysics today.

Digression — An Alternative

The spiral galaxy observations necessitate at least one of the following:

(1) There exists in galaxies large quantities of unseen matter which boosts the stars' velocities beyond what would be expected on the basis of the visible mass alone, or

(2) Newton's Laws do not apply to galaxies.

Option (1) leads to the dark matter hypothesis; option (2) leads to MOND.

The basic premise of MOND is that while Newton's laws have been extensively tested in high-acceleration environments (in the Solar System and on Earth), they have not been verified for objects with extremely low acceleration, such as stars in the outer parts of galaxies. This led Milgrom to postulate a new effective gravitational force law (sometimes referred to as "Milgrom's law") that relates the true acceleration of an object to the acceleration that would be predicted for it on the basis of Newtonian mechanics. This law, the keystone of MOND, is chosen to reproduce the Newtonian result at high acceleration but leads to different ("deep-MOND") behavior at low acceleration:

$$F_{\mathrm{N}} = m\,\mu\!\left(\frac{a}{a_0}\right) a$$

Here F_{N} is the Newtonian force, m is the object's (gravitational) mass, a is its acceleration, $\mu(x)$ is an as-yet unspecified function (called the *interpolating function*), and a_0 is a new fundamental constant which marks the transition between the Newtonian and deep-MOND regimes. Agreement with Newtonian mechanics requires

$$\mu(x) \longrightarrow 1 \qquad \text{for } x \gg 1$$

and consistency with astronomical observations requires

$$\mu(x) \longrightarrow x \qquad \text{for } x \ll 1$$

Beyond these limits, the interpolating function is not specified by the hypothesis, although it is possible to weakly constrain it empirically.

Two common choices are the "simple interpolating function":

$$\mu\left(\frac{a}{a_0}\right) = \frac{1}{1+\frac{a_0}{a}}$$

and the "standard interpolating function":

$$\mu\left(\frac{a}{a_0}\right) = \sqrt{\frac{1}{1+\left(\frac{a_0}{a}\right)^2}}$$

Thus, in the deep-MOND regime ($a \ll a_0$):

$$F_N = m\frac{a^2}{a_0}$$

Applying this to an object of mass *m* in circular orbit around a point mass *M* (a crude approximation for a star in the outer regions of a galaxy), we find:

$$\frac{GMm}{r^2} = m\frac{\left(\frac{v^2}{r}\right)^2}{a_0} \implies v^4 = GMa_0$$

that is, the star's rotation velocity is independent of *r*, its distance from the centre of the galaxy – the rotation curve is flat, as required.

By fitting his law to rotation curve data, Milgrom found $a_0 \approx 1.2 \times 10^{-10} ms^{-2}$ to be optimal. This simple law is sufficient to make predictions for a broad range of galactic phenomena.

Milgrom's law can be interpreted in two different ways:

One possibility is to treat it as a modification to Newton's Second Law, so that the force on an object is not proportional to the particle's acceleration *a* but rather to

$$\mu\left(\frac{a}{a_0}\right)a$$

in this case, the modified dynamics would apply not only to gravitational phenomena, but also to those generated by other forces.

Alternatively, Milgrom's law can be viewed as leaving Newton's Second Law intact and instead modifying the inverse-square law of gravity, so that the true gravitational force on an object of mass *m* due to another of mass *M* is roughly of the form

$$\frac{GMm}{\mu\left(\frac{a}{a_0}\right)r^2}$$

In this interpretation, Milgrom's modification would apply exclusively to gravitational phenomena.

Back to standard Classical Mechanics….

EXAMPLE 5.3

Consider a thin uniform circular ring of radius a and mass M. A mass m is placed in the plane of the ring. Find a position of equilibrium and determine whether it is stable.

Solution. From symmetry, we might believe that the mass m placed in the center of the ring (Figure 5-6) should be in equilibrium because it is uniformly surrounded by mass. Put mass m at a distance r' from the center of the ring, and place the x-axis along this direction.

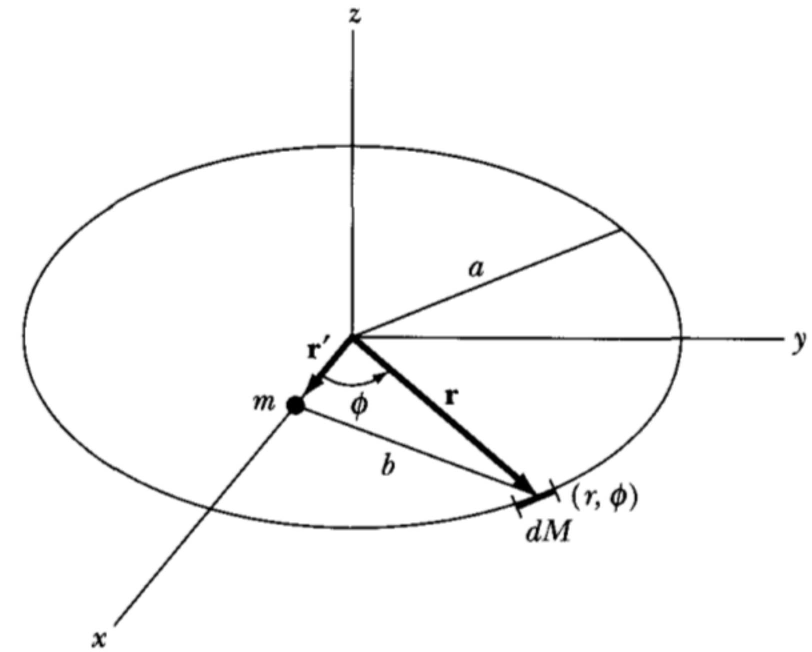


FIGURE 5-6 Example 5.3. The geometry of the point mass m and ring of mass M.

The potential is given by Equation 5.7 where $\rho = M/2\pi a$:

$$d\Phi = -G\frac{dM}{b} = -\frac{Ga\rho}{b}\,d\phi \tag{5.23}$$

where b is the distance between dM and m, and dM = $\rho a d\phi$. Let **r** and **r**' be the position vectors to dM and m, respectively.

$$\begin{aligned} b &= |\mathbf{r} - \mathbf{r}'| = |a\cos\phi\mathbf{e}_1 + a\sin\phi\mathbf{e}_2 - r'\mathbf{e}_1| \\ &= |(a\cos\phi - r')\mathbf{e}_1 + a\sin\phi\mathbf{e}_2| = [(a\cos\phi - r')^2 + a^2\sin^2\phi]^{1/2} \\ &= (a^2 + r'^2 - 2ar'\cos\phi)^{1/2} = a\left[1 + \left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^{1/2} \end{aligned} \tag{5.24}$$

Integrating Equation 5.23 gives

$$\begin{aligned} \Phi(r') &= -G\int\frac{dM}{b} = -\rho a G\int_0^{2\pi}\frac{d\phi}{b} \\ &= -\rho G\int_0^{2\pi}\frac{d\phi}{\left[1 + \left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^{1/2}} \end{aligned} \tag{5.25}$$

The integral in Equation 5.25 is difficult, so let us consider positions close to the equilibrium point, r' = 0. If r' << a, we can expand the denominator in Equation 5.25.

$$\left[1 + \left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^{-1/2} = 1 - \frac{1}{2}\left[\left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right] + \frac{3}{8}\left[\left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^2 + \cdots$$
$$= 1 + \frac{r'}{a}\cos\phi + \frac{1}{2}\left(\frac{r'}{a}\right)^2(3\cos^2\phi - 1) + \cdots \tag{5.26}$$

Equation 5.25 becomes

$$\Phi(r') = -\rho G\int_0^{2\pi}\left\{1 + \frac{r'}{a}\cos\phi + \frac{1}{2}\left(\frac{r'}{a}\right)^2(3\cos^2\phi - 1) + \cdots\right\}d\phi \tag{5.27}$$

which is easily integrated with the result

$$\Phi(r') = -\frac{MG}{a}\left[1 + \frac{1}{4}\left(\frac{r'}{a}\right)^2 + \cdots\right] \tag{5.28}$$

The potential energy U(r’) is from Equation 5.11, simply

$$U(r') = m\Phi(r') = -\frac{mMG}{a}\left[1 + \frac{1}{4}\left(\frac{r'}{a}\right)^2 + \cdots\right] \tag{5.29}$$

The position of equilibrium is found (from Equation 2.100) by

$$\frac{dU(r')}{dr'} = 0 = -\frac{mMG}{a}\frac{1}{2}\frac{r'}{a^2} + \cdots \tag{5.30}$$

so $r' = 0$ is an equilibrium point. We use Equation 2.103 to determine the stability:

$$\frac{d^2U(r')}{dr'^2} = -\frac{mMG}{2a^3} + \cdots < 0 \tag{5.31}$$

so the equilibrium point is unstable.

This last result is not obvious, because we might be led to believe that a small displacement from $r' = 0$ might still be returned to $r' = 0$ by the gravitational forces from all the mass in the ring surrounding it.

Poisson's Equation

It is useful to compare these properties of gravitational fields with some of the familiar results from electrostatics that were determined in the formulation of Maxwell's equations. Consider an arbitrary surface as in Figure 5-7 with a mass m placed somewhere inside. Similar to electric flux, let's find the gravitational flux Φ_m emanating from mass m through the arbitrary surface S.

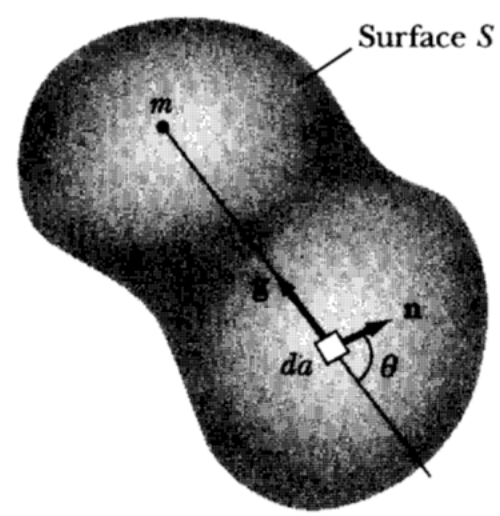


FIGURE 5-7 An arbitrary surface with a mass *m* placed inside. The unit vector **n** is normal to the surface at the differential area *da*.

$$\Phi_m = \int_S \mathbf{n} \cdot \mathbf{g}\, da \tag{5.32}$$

where the integral is over the surface S and the unit vector **n** is normal to the surface at the differential area da. If we substitute **g** from Equation 5.3 for the gravitational field vector for a body of mass m, we have for the scalar product **n** • **g**,

$$\mathbf{n}\cdot\mathbf{g} = -Gm\frac{\cos\theta}{r^2}$$

where θ is the angle between **n** and **g**. We substitute this into Equation 5.32 and obtain

$$\Phi_m = -Gm\int_S \frac{\cos\theta}{r^2}da$$

The integral is over the solid angle of the arbitrary surface and has the value 4π steradians, which gives for the mass flux

$$\Phi_m = \int_S \mathbf{n}\cdot\mathbf{g}\,da = -4\pi Gm \tag{5.33}$$

Note that it is immaterial where the mass is located inside the surface S. We can generalize this result for many masses m_i inside the surface S by summing over the masses.

$$\int_S \mathbf{n}\cdot\mathbf{g}\,da = -4\pi G\sum_i m_i \tag{5.34}$$

If we change to a continuous mass distribution within surface S, we have

$$\int_S \mathbf{n}\cdot\mathbf{g}\,da = -4\pi G\int_V \rho dv \tag{5.35}$$

where the integral on the right-hand side is over the volume V enclosed by S, ρ is the mass density, and dv is the differential volume. We use Gauss's divergence theorem to rewrite this result. Gauss's divergence theorem, Equation 1.130 where d**a** = **n** da, is

$$\int_S \mathbf{n}\cdot\mathbf{g}\, da = \int_V \nabla\cdot\mathbf{g}\, dv \tag{5.36}$$

If we set the right-hand sides of Equations 5.35 and 5.36 equal, we have

$$\int_V (-4\pi G)\rho dv = \int_V \nabla\cdot\mathbf{g}\, dv$$

and because the surface S, and its volume V, is completely arbitrary, the two integrands must be equal.

$$\nabla\cdot\mathbf{g} = -4\pi G\rho \tag{5.37}$$

This result is similar to the differential form of Gauss's law for electric field, $\nabla\cdot\mathbf{E} = \rho/\varepsilon$, where ρ in this case is the charge density.

We insert $\mathbf{g} = -\nabla\Phi$ from Equation 5.5 into the left-hand side of Equation 5.37 and obtain

$$\nabla\cdot\mathbf{g} = -\nabla\cdot\nabla\Phi = -\nabla^2\Phi$$

Equation 5.37 becomes

$$\nabla^2\Phi = 4\pi G\rho$$

which is known as *Poisson's equation* and is useful in a number of potential theory applications. When the right-hand side of Equation 5.38 is zero, the result $\nabla^2\Phi = 0$ is an even better known equation called *Laplace's equation*. Poisson's equation is useful in developing Green's functions, whereas we often encounter Laplace's equation when dealing with various coordinate systems.

5.3 Lines of Force and Equipotential Surfaces

Let us consider a mass that gives rise to a gravitational field that can be described by a field vector **g**. Let us draw a line outward from the surface of the mass such that the direction of the line at every point is the same as the direction of **g** at that point. This line will extend from the surface of the mass to infinity. Such a line is called a **line of force**.

By drawing similar lines from every small increment of surface area of the mass, we can indicate the direction of the force field at any arbitrary point in space. The lines of force for a single point mass are all straight lines extending from the mass to infinity. Defined in this way, the lines of force are related only to the *direction* of the force field at any point. We may consider, however, that the *density* of such lines - that is, the number of lines passing through a unit area oriented perpendicular to the lines—is proportional to the *magnitude* of the force at that area. The lines-of-force picture is thus a convenient way to visualize both the magnitude and the direction (i.e., the *vector* property) of the field.

The potential function is defined at every point in space (except at the position of a point mass). Therefore, the equation

$$\Phi = \Phi(x_1, x_2, x_3) = \text{constant} \tag{5.39}$$

defines a surface on which the potential is constant. Such a surface is called an equipotential surface. The field vector **g** is equal to the gradient of Φ, so **g** can have no component along an equipotential surface. It therefore follows that every line of force must be normal to every equipotential surface. Thus, the field does no work on a body moving along an equipotential surface. Because the potential function is single valued, no two equipotential surfaces can intersect or touch. The surfaces of equal potential that surround a single, isolated point mass (or any spherically symmetric mass) are all spheres.

Consider two point masses M that are separated by a certain distance. If r_1 is the distance from one mass to some point in space and if r_2 is the distance from the other mass to the same point, then

$$\Phi = -GM\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \text{constant} \tag{5.40}$$

defines the equipotential surfaces. Several of these surfaces are shown in Figure 5-8 for this two-particle system. In three dimensions, the surfaces are generated by rotating this diagram around the line connecting the two masses.

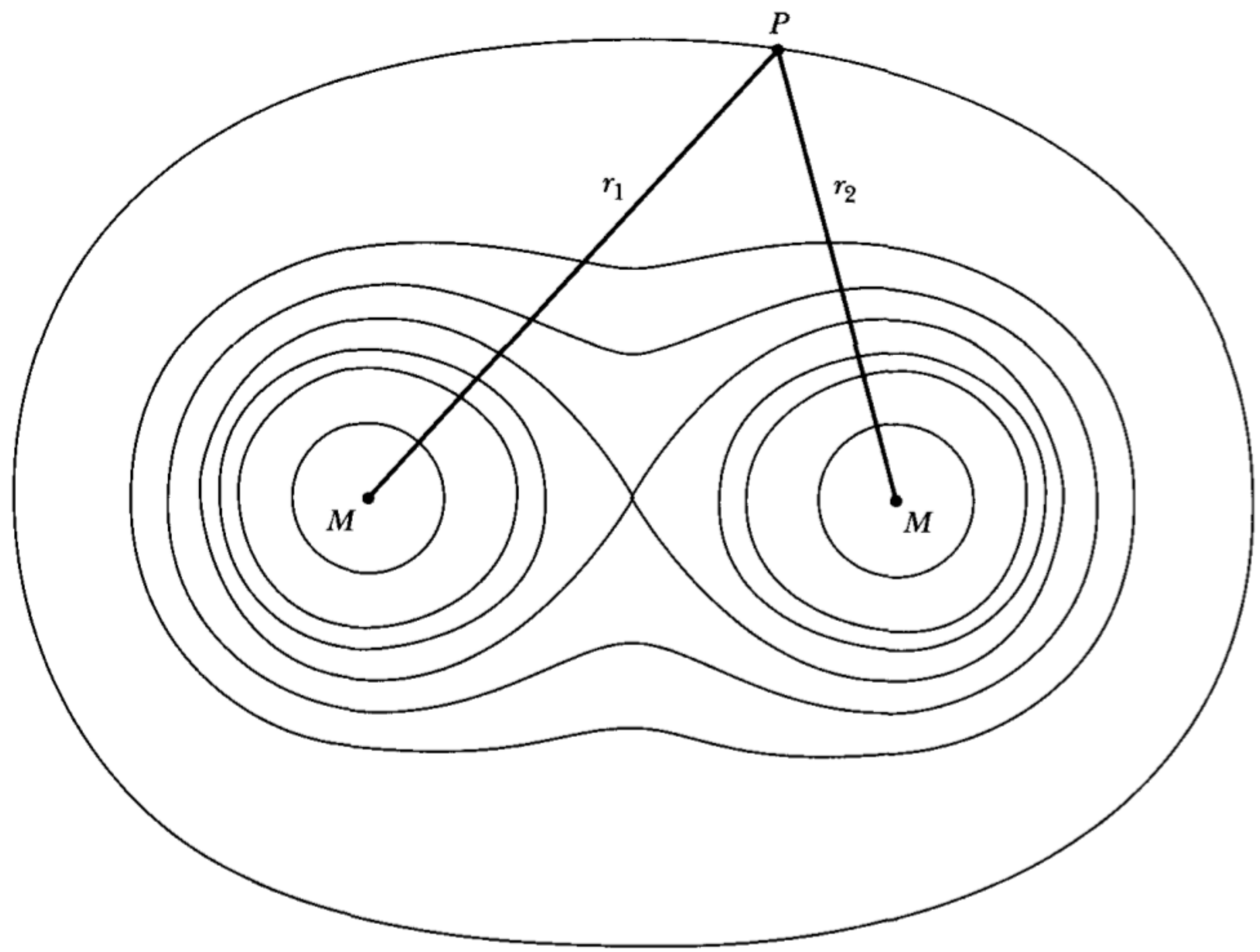


FIGURE 5-8 The equipotential surfaces due to two point masses *M*.

5.4 When Is the Potential Concept Useful?

The use of potentials to describe the effects of "action-at-a-distance" forces is an extremely important and powerful technique. We should not, however, lose sight of the fact that the ultimate justification for using a potential is to provide a convenient means of calculating the force on a body (or the energy for the body in the field) - for it is the force (and energy) and not the potential that is the physically meaningful quantity.

Thus, in some problems, it may be easier to calculate the force directly, rather than computing a potential and then taking the gradient. The advantage of using the potential method is that the potential is a scalar quantity: We need not deal with the added complication of sorting out the components of a vector until the gradient operation is performed. In direct calculations of the force, the components must be carried through the entire computation. Some skill, then, is necessary in choosing the particular approach to use. For example, if a problem has a particular symmetry that, from physical considerations, allows us to determine that the force has a certain direction, then the choice of that direction as one of the coordinate directions reduces the vector calculation to a simple scalar calculation. In such a case, the direct calculation of the force may be sufficiently straightforward to obviate the necessity of using the potential method. Every problem requiring a force must be examined to discover the easiest method of computation.

EXAMPLE 5.4

Consider a thin uniform disk of mass M and radius a. Find the force on a mass m located along the axis of the disk.

Solution. We solve this problem by using both the potential and direct force approaches. Consider Figure 5.9.

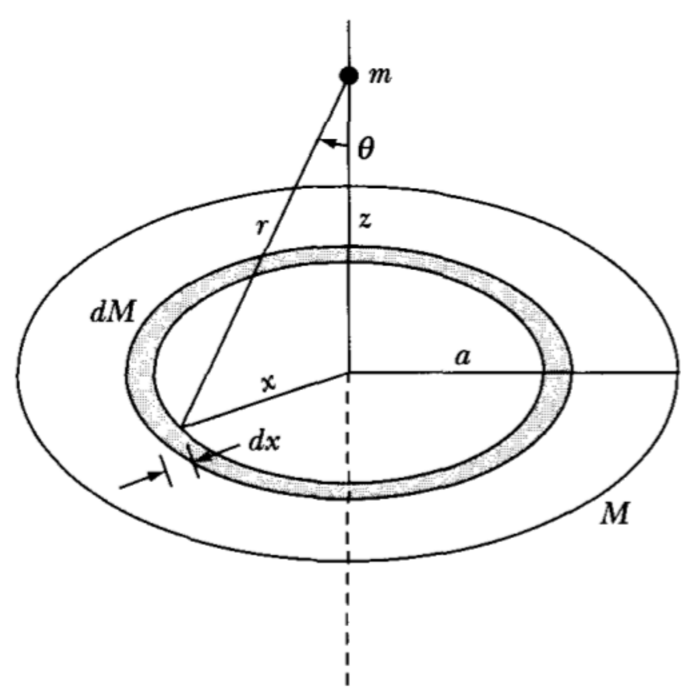


FIGURE 5-9 Example 5.4. We use the geometry shown here to find the gravitational force on a point mass *m* due to a thin uniform disk of mass *M*.

The differential potential dΦ at a distance z is given by

$$d\Phi = -G\frac{dM}{r} \tag{5.41}$$

The differential mass dM is a thin ring of width dx, because we have azimuthal symmetry.

$$dM = \rho dA = \rho 2\pi x\, dx \tag{5.42}$$

$$\begin{aligned} d\Phi &= -2\pi\rho G\frac{x\,dx}{r} = -2\pi\rho G\frac{x\,dx}{(x^2+z^2)^{1/2}} \\ \Phi(z) &= -\pi\rho G\int_0^a \frac{2x\,dx}{(x^2+z^2)^{1/2}} = -2\pi\rho G(x^2+z^2)^{1/2}\Big|_0^a \\ &= -2\pi\rho G[(a^2+z^2)^{1/2} - z] \end{aligned} \tag{5.43}$$

We find the force from

$$\mathbf{F} = -\nabla U = -m\nabla\Phi \tag{5.44}$$

From symmetry, we have only a force in the z direction

$$F_z = -m\frac{\partial\Phi(z)}{\partial z} = +2\pi m\rho G\left[\frac{z}{(a^2+z^2)^{1/2}} - 1\right] \tag{5.45}$$

In our second method, we compute the force directly using Equation 4.2:

$$d\mathbf{F} = -Gm\frac{dM'}{r^2}\mathbf{e}_r \tag{5.46}$$

where dM refers to the mass of a small differential area more like a square than a thin ring.

The vectors complicate matters. How can symmetry help? For every small dM′ on one side of the thin ring of width dx, another dM′ exists on the other side that exactly cancels the horizontal component of d**F** on m. Similarly, all horizontal components cancel, and we need only consider the vertical component of d**F** along z.

$$dF_z = \cos\theta \, |d\mathbf{F}| = -mG\frac{\cos\theta \, dM'}{r^2}$$

and, because $\cos\theta = z/r$,

$$dF_z = -mG\frac{z \, dM'}{r^3}$$

Now we integrate over the mass $dM' = \rho 2\pi x \, dx$ around the ring and obtain

$$dF_z = -mG\rho\frac{2\pi xz \, dx}{r^3}$$

and

$$\begin{aligned} F_z &= -\pi m\rho Gz \int_0^a \frac{2x \, dx}{(z^2 + x^2)^{3/2}} \\ &= -\pi m\rho Gz \left[\frac{-2}{(z^2 + x^2)^{1/2}}\right]\Bigg|_0^a \\ &= 2\pi m\rho G\left[\frac{z}{(a^2 + z^2)^{1/2}} - 1\right] \end{aligned} \tag{5.47}$$

which is identical to Equation 5.45. Notice that the value of F_z is negative, indicating that the force is downward in Figure 5-9 and attractive.

5.5 Ocean Tides

The ocean tides have long been of interest to humans. Galileo tried unsuccessfully to explain ocean tides but could not account for the timing of the approximately two high tides each day. Newton finally gave an adequate explanation. The tides are caused by the gravitational attraction of the ocean to both the Moon and the Sun, but there are several complicating factors.

The calculation is complicated by the fact that the surface of Earth is not an inertial system. Earth and Moon rotate about their center of mass (and move about the Sun), so we may regard the water nearest the Moon as being pulled away from Earth, and Earth as being pulled away from the water farthest from the Moon. However, Earth rotates while the Moon rotates about Earth. Let's first consider only the effect of the Moon, adding the effect of the Sun later. We will assume a simple model whereby Earth's surface is completely covered with water, and we shall add the effect of Earth's rotation at an appropriate time. We set up an inertial frame of reference x'y'z' as shown in Figure 5.10a.

We let M_m be the mass of the Moon, r the radius of a circular Earth, and D the distance from the center of the Moon to the center of Earth. We consider the effect of both the Moon's and Earth's gravitational attraction on a small mass m placed on the surface of Earth. As displayed in Figure 5-10a, the position vector of the mass m from the Moon is $\mathbf{R}$, from the center of Earth is $\mathbf{r}$, and from our inertial system $\mathbf{r}'_m$. The position vector from the inertial system to the center of Earth is $\mathbf{r}'_E$. As measured from the inertial system, the force on m, due to the earth and the Moon, is

$$m\ddot{\mathbf{r}}'_m = -\frac{GmM_E}{r^2}\mathbf{e}_r - \frac{GmM_m}{R^2}\mathbf{e}_R \tag{5.48}$$

Similarly, the force on the center of mass of Earth caused by the Moon is

$$M_E\ddot{\mathbf{r}}'_E = -\frac{GM_EM_m}{D^2}\mathbf{e}_D \tag{5.49}$$

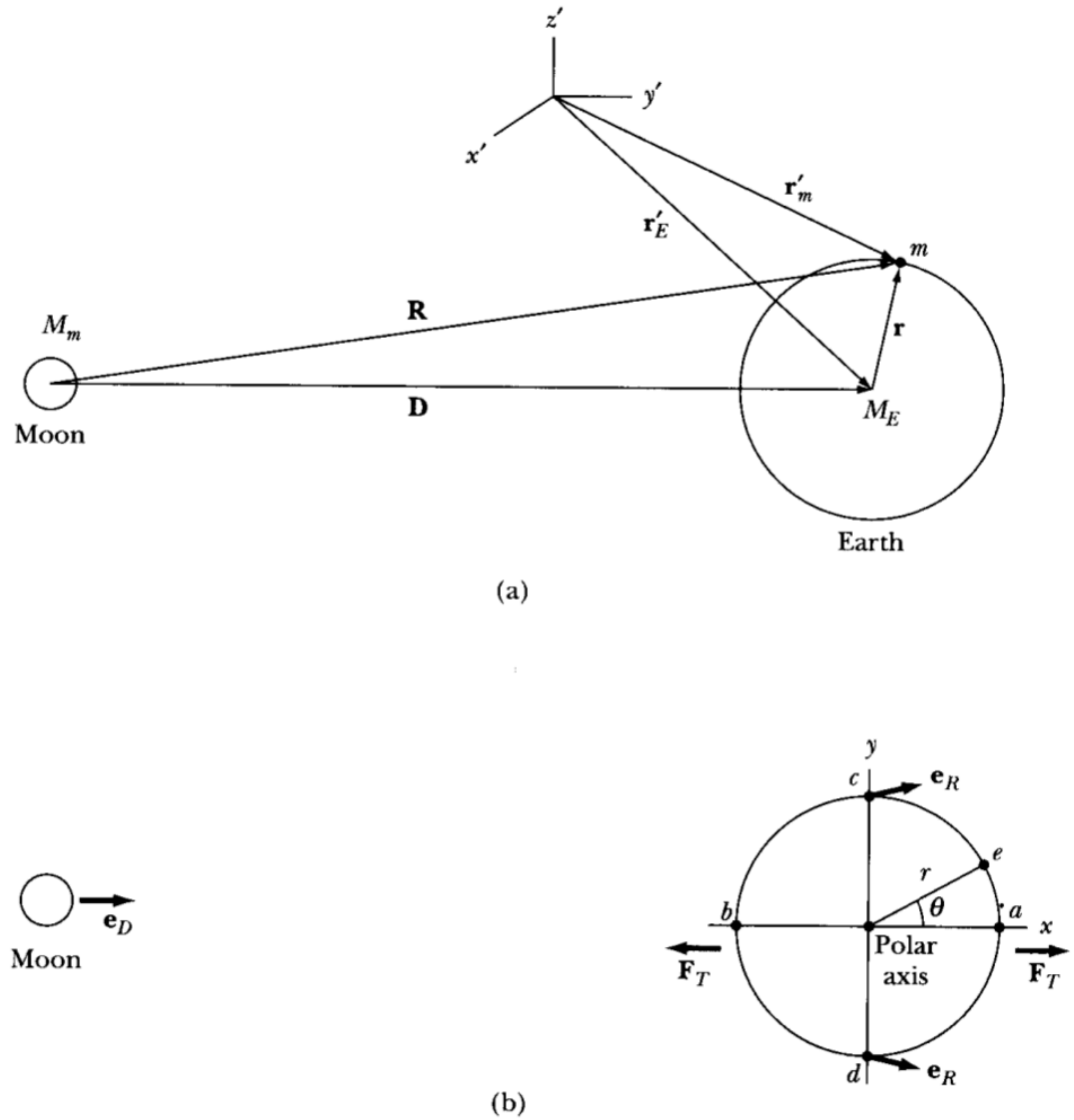


FIGURE 5-10 (a) Geometry to find ocean tides on Earth due to the Moon. (b) Polar view with the polar axis along the z-axis.

We want to find the acceleration $\ddot{r}$ as measured in the noninertial system placed at the center of Earth. Therefore, we want

$$
\begin{aligned}
\ddot{\mathbf{r}} = \ddot{\mathbf{r}}'_m - \ddot{\mathbf{r}}'_E &= \frac{m\ddot{\mathbf{r}}'_m}{m} - \frac{M_E\ddot{\mathbf{r}}'_E}{M_E} \\
&= -\frac{GM_E}{r^2}\mathbf{e}_r - \frac{GM_m}{R^2}\mathbf{e}_R + \frac{GM_m}{D^2}\mathbf{e}_D \\
&= -\frac{GM_E}{r^2}\mathbf{e}_r - GM_m\left(\frac{\mathbf{e}_R}{R^2} - \frac{\mathbf{e}_D}{D^2}\right)
\end{aligned}
\tag{5.50}
$$

The first part is due to Earth, and the second part is the acceleration from the **tidal** force, which is responsible for producing the ocean tides. It is due to the difference between the Moon's gravitational pull at the center of Earth and on Earth's surface.

We next find the effect of the tidal force at various points on Earth as noted in Figure 5-10b. We show a polar view of Earth with the polar axis along the z-axis. The tidal force $\mathbf{F}_T$ on the mass m on Earth's surface is

$$\mathbf{F}_T = -GmM_m\left(\frac{\mathbf{e}_R}{R^2} - \frac{\mathbf{e}_D}{D^2}\right) \tag{5.51}$$

where we have used only the second part of Equation 5.50. We look first at point a, the farthest point on Earth from the Moon. Both unit vectors $\mathbf{e}_R$ and $\mathbf{e}_D$ are pointing in the same direction away from the Moon along the x-axis. Because $R > D$, the second term in Equation 5.51 predominates, and the tidal force is along the +x-axis as shown in Figure 5-l0b. For point b, $R < D$ and the tidal force has approximately the same magnitude as at point a because $R/D \ll 1$, but is along the -x-axis. The magnitude of the tidal force along the x-axis, F_{Tx}, is

$$F_{Tx} = -GmM_m\left(\frac{1}{R^2} - \frac{1}{D^2}\right) = -GmM_m\left(\frac{1}{(D+r)^2} - \frac{1}{D^2}\right)$$

$$= -\frac{GmM_m}{D^2}\left(\frac{1}{\left(1+\frac{r}{D}\right)^2} - 1\right)$$

We expand the first term in brackets using the $(1 + x)^{-2}$ expansion in Equation D.9.

$$F_{Tx} = -\frac{GmM_m}{D^2}\left[1 - 2\frac{r}{D} + 3\left(\frac{r}{D}\right)^2 - \cdots - 1\right] = +\frac{2GmM_m r}{D^3} \tag{5.52}$$

where we have kept only the largest nonzero term in the expansion, because r/D = 0.02.

For point c, the unit vector $\mathbf{e}_R$ (Figure 5-l0b) is not quite exactly along $\mathbf{e}_D$, but the x-axis components approximately cancel, because $R \simeq D$ and the x-components of $\mathbf{e}_R$ and $\mathbf{e}_D$ are similar, There will be a small component of $\mathbf{e}_R$ along the y-axis. We approximate the y-component of $\mathbf{e}_R$ by (r/D))$\mathbf{j}$, and the tidal force at point c, call it $\mathbf{F}_{Ty}$, is along the y-axis and has the magnitude

$$F_{Ty} = -GmM_m\left(\frac{1}{D^2}\frac{r}{D}\right) = -\frac{GmM_m r}{D^3} \tag{5.53}$$

Note that this force is along the -y-axis toward the center of Earth at point c. We find similarly at point D the same magnitude, but the component of $\mathbf{e}_R$ will be along the -y-axis, so the force itself, with the sign of Equation 5.53, will be along the +y-axis toward the center of Earth. We indicate the tidal forces at points a, b, c, and don Figure 5-11 a.

We determine the force at an arbitrary point e by noting that the x- and y-components of the tidal force can be found by substituting x and y for r in F_{Tx} and F_{Ty}, respectively, in Equations 5.52 and 5.53.

$$F_{Tx} = \frac{2GmM_m x}{D^3}$$

$$F_{Ty} = -\frac{GmM_m y}{D^3}$$

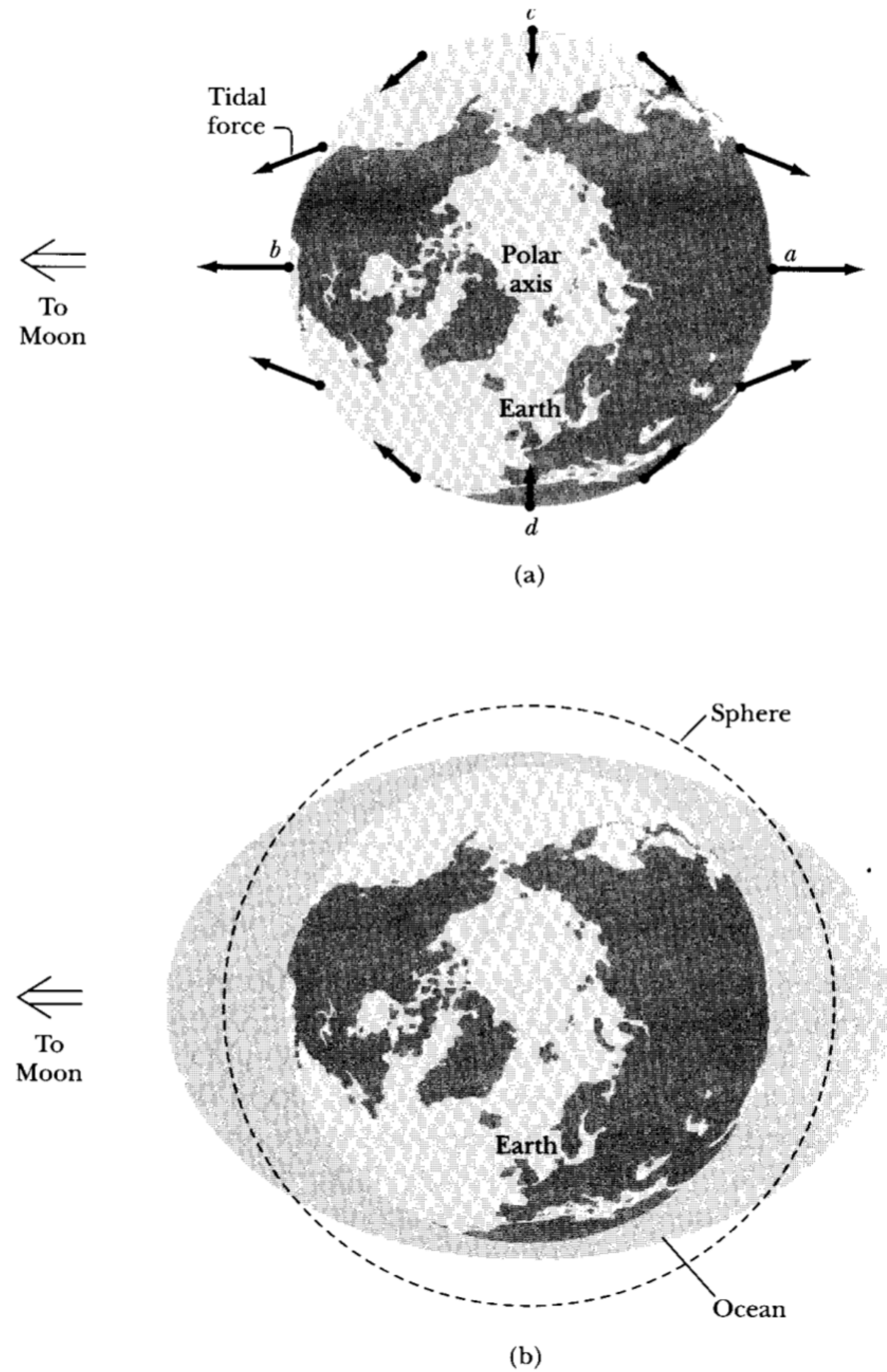


FIGURE 5-11 (a) The tidal forces are shown at various places on Earth's surface including the points *a*, *b*, *c*, and *d* of Figure 5-10. (b) An exaggerated view of Earth's ocean tides.

Then at an arbitrary point such as e, we let $x = r\cos\theta$ and $y = r\sin\theta$, so we have

$$F_{Tx} = \frac{2GmM_m r\cos\theta}{D^3} \tag{5.54a}$$

$$F_{Ty} = -\frac{GmM_m r\sin\theta}{D^3} \tag{5.54b}$$

Equations 5.54a and b give the tidal force around Earth for all angles θ. Note that they give the correct result at points a, b, c, and d.

Figure 5-11a gives a representation of the tidal forces. For our simple model, these forces lead to the water along the y-axis being more shallow than along the x-axis. We show an exaggerated result in Figure 5-11b. As Earth makes a revolution about its own axis every 24 hours, we will observe two high tides a day.

A quick calculation shows that the Sun's gravitational attraction is about 175 times stronger than the Moon's on Earth's surface, so we would expect tidal forces from the Sun as well. The tidal force calculation is similar to the one we have just performed for the Moon. The result is that the tidal force due to the Sun is 0.46 that of the Moon, a sizable effect. Despite the stronger attraction due to the Sun, the gravitational force gradient over the surface of Earth is much smaller, because of the much larger distance to the Sun

EXAMPLE 5.5

Calculate the maximum height change in the ocean tides caused by the Moon.

Solution. We continue to use our simple model of the ocean surrounding Earth. Newton proposed a solution to this calculation by imagining that two wells be dug, one along the direction of high tide (our x-axis) and one along the direction of low tide (our y-axis). If the tidal height change we want to determine is h, then the difference in potential energy of mass m due to the height difference is mgh.

Let's calculate the difference in work if we move the mass m from point c in Figure 5-12 to the center of Earth and then to point a.

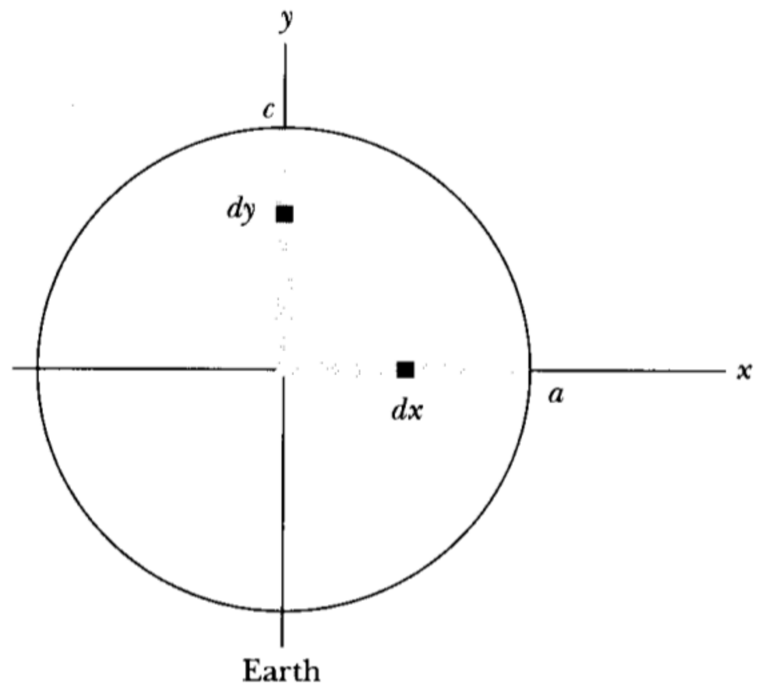


FIGURE 5-12 Example 5.5. We calculate the work done to move a point mass *m* from point *c* to the center of Earth and then to point *a*.

This work W done by gravity must equal the potential energy change mgh. The work W is

$$W = \int_{r+\delta_1}^{0} F_{Ty}\,dy + \int_{0}^{r+\delta_2} F_{Tx}\,dx$$

where we use the tidal forces F_{Ty} and F_{Tx} of Equations 5.54. The small distances δ_1 and δ_2 are to account for the small variations from a spherical Earth, but these values are so small they can be henceforth neglected. The value for W becomes

$$W = \frac{GmM_m}{D^3}\left[\int_r^0 (-y)\,dy + \int_0^r 2x\,dx\right]$$
$$= \frac{GmM_m}{D^3}\left(\frac{r^2}{2} + r^2\right) = \frac{3GmM_m r^2}{2D^3}$$

Because this work is equal to mgh, we have

$$mgh = \frac{3GmM_m r^2}{2D^3}$$

$$h = \frac{3GM_m r^2}{2gD^3} \tag{5.55}$$

Note that the mass m cancels, and the value of h does not depend on m. Nor does it depend on the substance, so to the extent Earth is plastic, similar tidal effects should be (and are) observed for the surface land. If we insert the known values of the constants into Equation 5.55, we find

$$h = \frac{3(6.67 \times 10^{-11} \text{m}^3/\text{kg}\cdot\text{s}^2)(7.350 \times 10^{22}\text{kg})(6.37 \times 10^6\text{m})^2}{2(9.80\text{ m/s}^2)(3.84 \times 10^8\text{m})^3} = 0.54 \text{ m}$$

The highest tides (called *spring* tides) occur when Earth, the Moon, and the Sun are lined up (new moon and full moon), and the smallest tides (called *neap* tides) occur for the first and third quarters of the Moon when the Sun and Moon are at right angles to each other, partially cancelling their effects. The maximum tide, which occurs every 2 weeks, should be 1.46h = 0.83 m for the spring tides.

An observer who has spent much time near the ocean has noticed that typical ocean shore tides are greater than those calculated in Example 5.5. Several other effects come into play. Earth is not covered completely with water, and the continents play a significant role, especially the shelfs and narrow estuaries. Local effects can be dramatic, leading to tidal changes of several meters. The tides in midocean, however, are similar to what we have calculated. Resonances can affect the natural oscillation of the bodies of water and cause tidal changes.

Tidal friction between water and Earth leads to a significant amount of energy loss on Earth. Earth is not rigid, and it is also distorted by tidal forces.

In addition to the effects just discussed, remember that as Earth rotates, the Moon is also orbiting Earth. This leads to the result that there are not quite exactly two high tides per day, because they occur once every 12 h and 26 min. The plane of the moon's orbit about Earth is also not perpendicular to Earth's rotation axis. This causes one high tide each day to be slightly higher than the other. The tidal friction between water and land mentioned previously also results in Earth "dragging" the ocean with it as Earth rotates. This causes the high tides to be not quite along the Earth-Moon axis, but rather several degrees apart as shown in Figure 5-13.

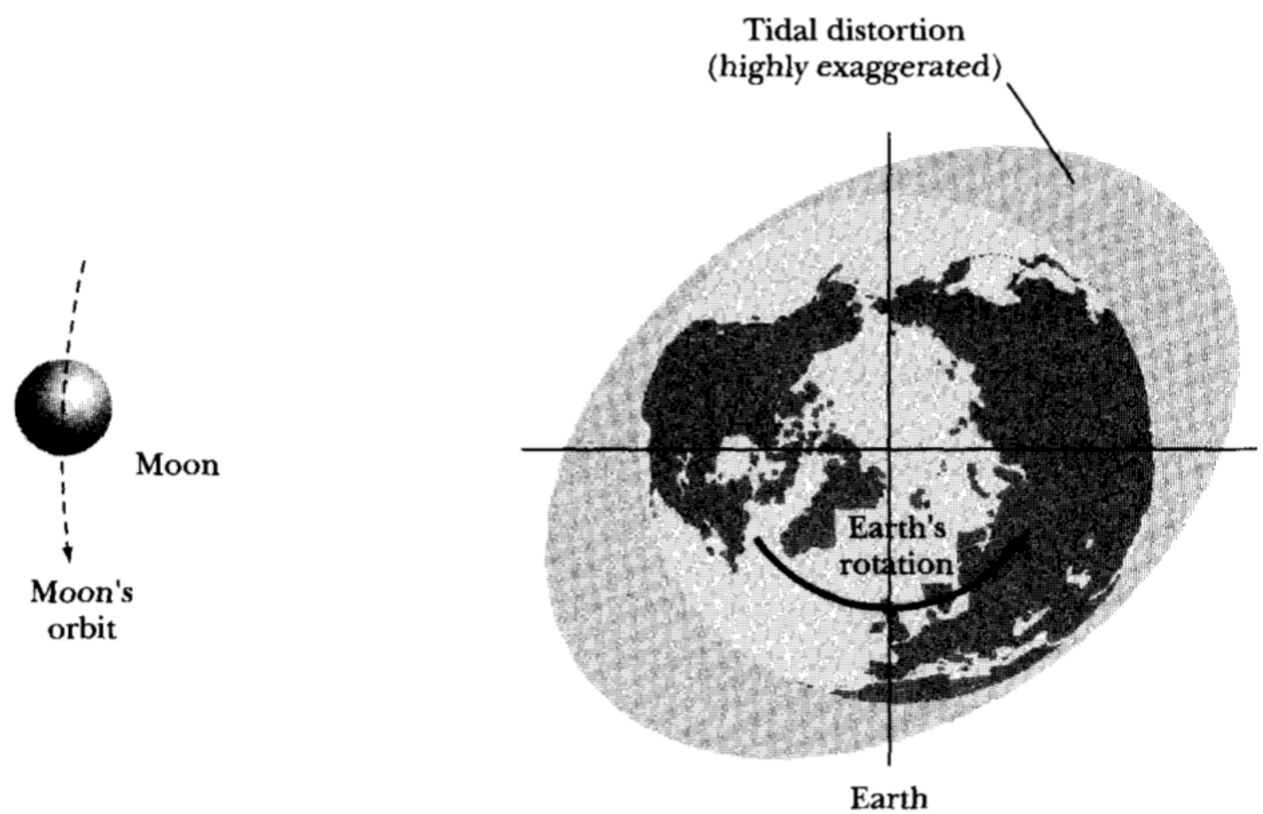


FIGURE 5-13 Some effects cause the high tides to not be exactly along the Earth-Moon axis.

CHAPTER 8 - Central-Force Motion

8.1 Introduction

The motion of a system consisting of two bodies affected by a force directed along the line connecting the centers of the two bodies (i.e., a *central force*) is an extremely important physical problem - one we can solve completely. The importance of such a problem lies in large measure in two quite different realms of physics: the motion of celestial bodies - planets, moons, comets, double stars, and the like - and certain two-body nuclear interactions, such as the scattering of a particles by nuclei. In the pre-quantum mechanics days, physicists also described the hydrogen atom in terms of a classical two-body central force. Although such a description is still useful in a qualitative sense, the quantum-theoretical approach must be used for a detailed description. In addition to some general considerations regarding motion in central-force fields, we discuss in this and the following chapter several of the problems of two bodies encountered in celestial mechanics and in nuclear and particle physics.

8.2 Reduced Mass

Describing a system consisting of two particles requires the specification of six quantities; for example, the three components of each of the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ for the particles. The orientation of the particles is assumed to be unimportant; that is, they are spherically symmetric (or are point particles). Alternatively, we may choose the three components of the center-of-mass vector $\mathbf{R}$ and the three components of $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ (see Figure 8-la).

Here, we restrict our attention to systems without frictional losses and for which the potential energy is a function only of $r = |\mathbf{r}_1 - \mathbf{r}_2|$. The Lagrangian for such a system may be written as

$$L = \frac{1}{2} m_1 |\dot{\mathbf{r}}_1|^2 + \frac{1}{2} m_2 |\dot{\mathbf{r}}_2|^2 - U(r) \tag{8.1}$$

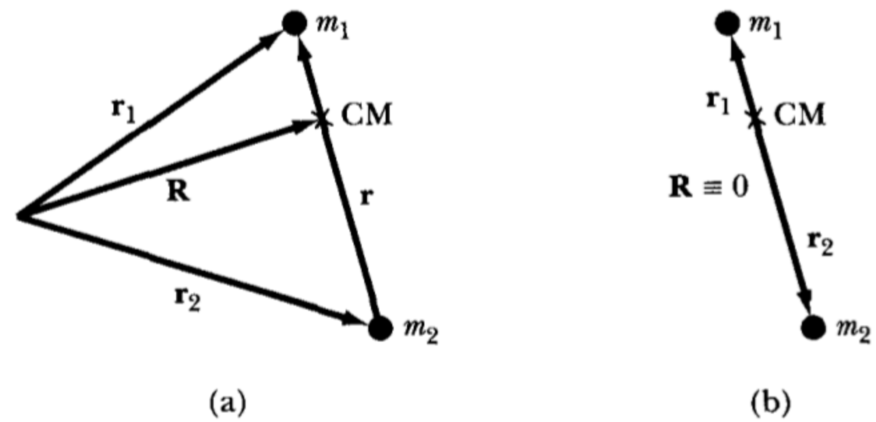


FIGURE 8-1 Two methods to describe the position of two particles. (a) From an arbitrary coordinate system origin, and (b) from the center of mass. The position vectors are $\mathbf{r}_1$ and $\mathbf{r}_2$, the center-of-mass vector is **R**, and the relative vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$.

Because translational motion of the system as a whole is uninteresting from the standpoint of the particle orbits with respect to one another, we may choose the origin for the coordinate system to be the particles' center of mass - that is, $\mathbf{R} = 0$ (see Figure 8-lb). Then (see Section 9.2)

$$m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = 0 \tag{8.2}$$

This equation, combined with $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, yields

$$\left.\begin{aligned} \mathbf{r}_1 &= \frac{m_2}{m_1 + m_2}\mathbf{r} \\ \mathbf{r}_2 &= -\frac{m_1}{m_1 + m_2}\mathbf{r} \end{aligned}\right\} \tag{8.3}$$

Substituting Equation 8.3 into the expression for the Lagrangian gives

$$\boxed{L = \frac{1}{2}\mu|\dot{\mathbf{r}}|^2 - U(r)} \tag{8.4}$$

where μ , is the **reduced mass**,

$$\boxed{\mu \equiv \frac{m_1 m_2}{m_1 + m_2}} \tag{8.5}$$

We have therefore formally reduced the problem of the motion of two bodies to an equivalent one-body problem in which we must determine only the motion of a "particle" of mass μ , in the central field described by the potential function U(r).

Once we obtain the solution for **r**(t) by applying the Lagrange equations to Equation 8.4, we can find the individual motions of the particles, $\mathbf{r}_1$(t) and r_2(t), by using Equation 8.3. This latter step is not necessary if only the orbits relative to one another are required.

8.3 Conservation Theorems - First Integrals of the Motion

The system we wish to discuss consists of a particle of mass μ moving in a central-force field described by the potential function U(r). Because the potential energy depends only on the distance of the particle from the force center and not on the orientation, the system possesses **spherical symmetry**; that is, the system's rotation about any fixed axis through the center of force cannot affect the equations of motion. We have already shown (see Section 7.9) that under such conditions the angular momentum of the system is conserved:

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \text{constant} \tag{8.6}$$

From this relation, it should be clear that both the radius vector and the linear momentum vector of the particle lie always in a plane normal to the angular momentum vector L, which is fixed in space (see Figure 8-2). Therefore, we have only a two-dimensional problem, and the Lagrangian may then be conveniently expressed in plane polar coordinates:

$$\boxed{L = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2) - U(r)} \tag{8.7}$$

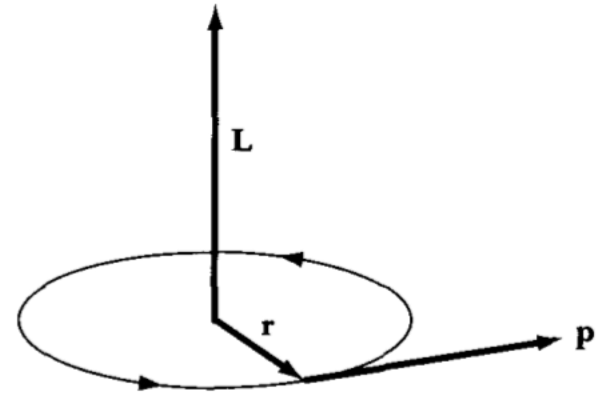


FIGURE 8-2 The motion of a particle of mass μ moving in a central-force field is described by the position vector **r**, linear momentum **p**, and constant angular momentum **L**.

Because the Lagrangian is cyclic in θ, the angular momentum conjugate to the coordinate θ is conserved:

$$\dot{p}_\theta = \frac{\partial L}{\partial \theta} = 0 = \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} \tag{8.8}$$

or

$$p_\theta \equiv \frac{\partial L}{\partial \dot{\theta}} = \mu r^2 \dot{\theta} = \text{constant} \tag{8.9}$$

The system's symmetry has therefore permitted us to integrate immediately one of the equations of motion. The quantity p_θ is a *first integral* of the motion, and we denote its constant value by the symbol l:

$$\boxed{l \equiv \mu r^2 \dot{\theta} = \text{constant}} \tag{8.10}$$

Note that l can be negative as well as positive. That l is constant has a simple geometric interpretation. Referring to Figure 8-3, we see that in describing the path **r**(t), the radius vector sweeps out an area $\frac{1}{2}r^2 d\theta$ in a time interval dt

$$dA = \frac{1}{2} r^2 d\theta \tag{8.11}$$

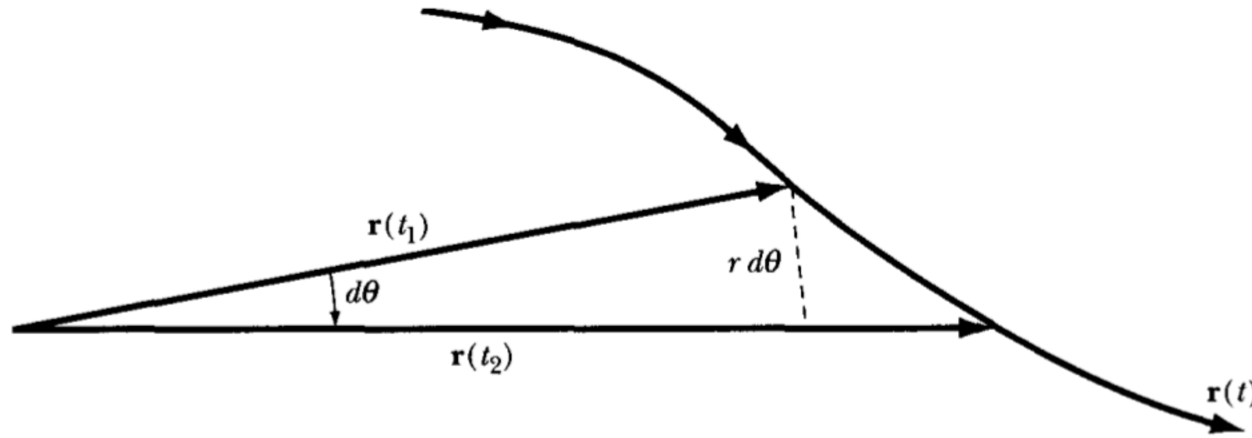


FIGURE 8-3 The path of a particle is described by $\mathbf{r}(t)$. The radius vector sweeps out an area $dA = \frac{1}{2}r^2 d\theta$ in a time interval dt.

On dividing by the time interval, the areal velocity is shown to be

$$\begin{aligned} \frac{dA}{dt} &= \frac{1}{2}r^2\frac{d\theta}{dt} = \frac{1}{2}r^2\dot{\theta} \\ &= \frac{l}{2\mu} = \text{constant} \end{aligned} \tag{8.12}$$

Thus, the areal velocity is constant in time. This result was obtained empirically by Kepler for planetary motion, and it is known as **Kepler's Second Law**. It is important to note that the conservation of the areal velocity is not limited to an inverse-square-law force (the case for planetary motion) but is a general result for central-force motion.

Because we have eliminated from consideration the uninteresting uniform motion of the system's center of mass, the conservation of linear momentum adds nothing new to the description of the motion. The conservation of energy is thus the only remaining first integral of the problem. The conservation of the total energy E is automatically ensured because we have limited the discussion to nondissipative systems. Thus,

$$T + U = E = \text{constant} \tag{8.13}$$

and

$$E = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2) + U(r)$$

or

$$\boxed{E = \frac{1}{2}\mu\dot{r}^2 + \frac{1}{2}\frac{l^2}{\mu r^2} + U(r)} \tag{8.14}$$

8.4 Equations of Motion

When U(r) is specified, Equation 8.14 completely describes the system, and the integration of this equation gives the general solution of the problem in terms of the parameters E and l. Solving Equation 8.14 for $\dot{r}$, we have

$$\dot{r} = \frac{dr}{dt} = \pm\sqrt{\frac{2}{\mu}(E - U) - \frac{l^2}{\mu^2 r^2}} \tag{8.15}$$

This equation can be solved for dt and integrated to yield the solution t = t{(r). An inversion of this result then gives the equation of motion in the standard form r = r(t). At present, however, we are interested in the equation of the path in terms of r and θ. We can write

$$d\theta = \frac{d\theta}{dt}\frac{dt}{dr}dr = \frac{\dot{\theta}}{\dot{r}}dr \tag{8.16}$$

Into this relation, we can substitute $\dot{\theta} = l/\mu r^2$ (Equation 8.10) and the expression for $\dot{r}$ from Equation 8.15. Integrating, we have

$$\theta(r) = \int \frac{\pm(l/r^2)\,dr}{\sqrt{2\mu\left(E - U - \dfrac{l^2}{2\mu r^2}\right)}} \tag{8.17}$$

Furthermore, because l is constant in time, $\dot{\theta}$ cannot change sign and therefore $\theta(t)$ must increase or decrease monotonically with time.

Although we have reduced the problem to the formal evaluation of an integral, the actual solution can be obtained only for certain specific forms of the force law. If the force is proportional to some power of the radial distance, $F(r) \propto r^n$, then the solution can be expressed in terms of elliptic integrals for certain integer and fractional values of n. Only for n = 1, -2, and -3 are the solutions expressible in terms of circular functions (sines and cosines). The case n = 1 is just that of the harmonic oscillator (see Chapter 3), and the case n = — 2 is the important inverse-square-law force treated in Sections 8.6 and 8.7. These two cases, n = 1, —2, are of prime importance in physical situations.

We have therefore solved the problem in a formal way by combining the equations that express the conservation of energy and angular momentum into a single result, which gives the equation of the orbit $\theta = \theta(r)$. We can also attack the problem using Lagrange's equation for the coordinate r:

$$\frac{\partial L}{\partial r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{r}} = 0$$

Using Equation 8.7 for L, we find

$$\mu(\ddot{r} - r\dot{\theta}^2) = -\frac{\partial U}{\partial r} = F(r) \tag{8.18}$$

Equation 8.18 can be cast in a form more suitable for certain types of calculations by making a simple change of variable:

$$u \equiv \frac{1}{r}$$

First, we compute

$$\frac{du}{d\theta} = -\frac{1}{r^2}\frac{dr}{d\theta} = -\frac{1}{r^2}\frac{dr}{dt}\frac{dt}{d\theta} = -\frac{1}{r^2}\frac{\dot{r}}{\dot{\theta}}$$

But from Equation 8.10, $\dot{\theta} = l/\mu r^2$, so

$$\frac{du}{d\theta} = -\frac{\mu}{l}\dot{r}$$

Next, we write

$$\frac{d^2u}{d\theta^2} = \frac{d}{d\theta}\left(-\frac{\mu}{l}\dot{r}\right) = \frac{dt}{d\theta}\frac{d}{dt}\left(-\frac{\mu}{l}\dot{r}\right) = -\frac{\mu}{l\dot{\theta}}\ddot{r}$$

and with the same substitution for $\dot{\theta}$, we have

$$\frac{d^2u}{d\theta^2} = -\frac{\mu^2}{l^2}r^2\ddot{r}$$

Therefore, solving for $\ddot{r}$ and $r\dot{\theta}^2$ in terms of u, we find

$$\left.\begin{aligned} \ddot{r} &= -\frac{l^2}{\mu^2}u^2\frac{d^2u}{d\theta^2} \\ r\dot{\theta}^2 &= \frac{l^2}{\mu^2}u^3 \end{aligned}\right\} \tag{8.19}$$

Substituting Equation 8.19 into Equation 8.18, we obtain the transformed equation of motion:

$$\frac{d^2u}{d\theta^2} + u = -\frac{\mu}{l^2}\frac{1}{u^2}F(1/u) \tag{8.20}$$

which we may also write as

$$\boxed{\frac{d^2}{d\theta^2}\left(\frac{1}{r}\right) + \frac{1}{r} = -\frac{\mu r^2}{l^2}F(r)} \tag{8.21}$$

This form of the equation of motion is particularly useful if we wish to find the force law that gives a particular known orbit $r = r(\theta)$.

EXAMPLE 8.1

Find the force law for a central-force field that allows a particle to move in a logarithmic spiral orbit given by $r = ke^{\alpha\theta}$, where k and α are constants.

Solution. We use Equation 8.21 to determine the force law F(r). First, we determine

$$\frac{d}{d\theta}\left(\frac{1}{r}\right) = \frac{d}{d\theta}\left(\frac{e^{-\alpha\theta}}{k}\right) = \frac{-\alpha e^{-\alpha\theta}}{k}$$

$$\frac{d^2}{d\theta^2}\left(\frac{1}{r}\right) = \frac{\alpha^2 e^{-\alpha\theta}}{k} = \frac{\alpha^2}{r}$$

From Equation 8.21, we now determine F(r).

$$F(r) = \frac{-l^2}{\mu r^2}\left(\frac{\alpha^2}{r} + \frac{1}{r}\right)$$

$$F(r) = \frac{-l^2}{\mu r^3}(\alpha^2 + 1) \tag{8.22}$$

Thus, the force law is an attractive inverse cube.

EXAMPLE 8.2

Determine r(t) and θ(t) for the problem in Example 8.1.

Solution. From Equation 8.10, we find

$$\dot{\theta} = \frac{l}{\mu r^2} = \frac{l}{\mu k^2 e^{2\alpha\theta}} \tag{8.23}$$

Rearranging Equation 8.23 gives

$$e^{2\alpha\theta}\,d\theta = \frac{l}{\mu k^2}\,dt$$

and integrating gives

$$\frac{e^{2\alpha\theta}}{2\alpha} = \frac{lt}{\mu k^2} + C'$$

where C is an integration constant. Multiplying by 2α and letting $C = 2\alpha C'$ gives

$$e^{2\alpha\theta} = \frac{2\alpha lt}{\mu k^2} + C \tag{8.24}$$

We solve for θ(t) by taking the natural logarithm of Equation 8.24:

$$\theta(t) = \frac{1}{2\alpha}\ln\left(\frac{2\alpha lt}{\mu k^2} + C\right) \tag{8.25}$$

We can similarly solve for r(t) by examining Equations 8.23 and 8.24:

$$\begin{aligned} \frac{r^2}{k^2} &= e^{2\alpha\theta} = \frac{2\alpha lt}{\mu k^2} + C \\ r(t) &= \left[\frac{2\alpha l}{\mu}t + k^2 C\right]^{1/2} \end{aligned} \tag{8.26}$$

The integration constant C and angular momentum l needed for Equations 8.25 and 8.26 are determined from the initial conditions.

EXAMPLE 8.3

What is the total energy of the orbit of the previous two examples?

Solution. The energy is found from Equation 8.14. In particular, we need $\dot{r}$ and U(r).

$$U(r) = -\int F\,dr = \frac{+l^2}{\mu}(\alpha^2 + 1)\int r^{-3}dr$$
$$U(r) = -\frac{l^2(\alpha^2 + 1)}{2\mu}\frac{1}{r^2} \tag{8.27}$$

where we have let U(∞) = 0.

We rewrite Equation 8.10 to determine $\dot{r}$:

$$\dot{\theta} = \frac{d\theta}{dt} = \frac{d\theta}{dr}\frac{dr}{dt} = \frac{1}{\mu r^2}$$
$$\dot{r} = \frac{dr}{d\theta}\frac{l}{\mu r^2} = \alpha k e^{\alpha\theta}\frac{l}{\mu r^2} = \frac{\alpha l}{\mu r} \tag{8.28}$$

Substituting Equations 8.27 and 8.28 into Equation 8.14 gives

$$E = \frac{1}{2}\mu\left(\frac{\alpha l}{\mu r}\right)^2 + \frac{l^2}{2\mu r^2} - \frac{l^2(\alpha^2 + 1)}{2\mu r^2} \tag{8.29}$$
$$E = 0$$

The total energy of the orbit is zero if U(r = ∞) = 0.

8.5 Orbits in a Central Field

The radial velocity of a particle moving in a central field is given by Equation 8.15. This equation indicates that $\dot{r}$ vanishes at the roots of the radical, that is, at points for which

$$E - U(r) - \frac{l^2}{2\mu r^2} = 0 \qquad (8.30)$$

The vanishing of $\dot{r}$ implies that a turning point in the motion has been reached (see Section 2.6). In general, Equation 8.30 possesses two roots: r_{max} and r_{min}. The motion of the particle is therefore confined to the annular region specified by $r_{max} \geq r \geq r_{min}$. Certain combinations of the potential function U(r) and the parameters E and l produce only a single root for Equation 8.30. In such a case, $\dot{r} = 0$ for all values of the time; hence, r = constant, and the orbit is circular.

If the motion of a particle in the potential U(r) is periodic, then the orbit is *closed*; that is, after a finite number of excursions between the radial limits r_{min} and r_{max} the motion exactly repeats itself. But if the orbit does not close on itself after a finite number of oscillations, the orbit is said to be *open* (Figure 8-4).

From Equation 8.17, we can compute the change in the angle θ that results from one complete transit of r from r_{min} to r_{max} and back to r_{min}. Because the motion is symmetric in time, this angular change is twice that which would result from the passage from r_{min} to r_{max}; thus

$$\Delta\theta = 2\int_{r_{min}}^{r_{max}} \frac{(l/r^2)\,dr}{\sqrt{2\mu\left(E - U - \dfrac{l^2}{2\mu r^2}\right)}} \qquad (8.31)$$

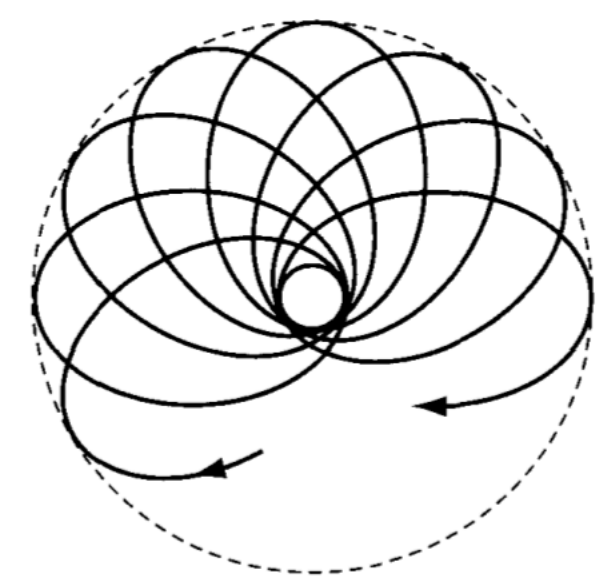


FIGURE 8-4 An orbit that does not close on itself after a finite number of oscillations is said to be *open*.

The path is closed only if $\Delta\theta$ is a rational fraction of 2π - that is, if , where a and b are integers. Under these conditions, after b periods the radius vector of the particle will have made a complete revolutions and will have returned to its original position. One can show that if the potential varies with some integer power of the radial distance, $\Delta\theta = 2\pi(a/b)$ then a closed noncircular path can result only if n = -2 or +1. The case n = -2 corresponds to an inverse-square-law force - for example, the gravitational or electrostatic force. The n = +1 case corresponds to the harmonic oscillator potential. For the two-dimensional case discussed in Section 3.4, we found that a closed path for the motion resulted if the ratio of the angular frequencies for the x and y motions were rational.

8.6 Centrifugal Energy and the Effective Potential

In the preceding expressions for $\dot{r}$, $\Delta\theta$, and so forth, a common term is the radical

$$\sqrt{E - U - \frac{l^2}{2\mu r^2}}$$

The last term in the radical has the dimensions of energy and, according to Equation 8.10, can also be written as

$$\frac{l^2}{2\mu r^2} = \frac{1}{2}\mu r^2\dot{\theta}^2$$

If we interpret this quantity as a "potential energy",

$$U_c \equiv \frac{l^2}{2\mu r^2} \tag{8.32}$$

then the "force" that must be associated with U_c is

$$F_c = -\frac{\partial U_c}{\partial r} = \frac{l^2}{\mu r^3} = \mu r\dot{\theta}^2 \tag{8.33}$$

This quantity is traditionally called the **centrifugal force**, although it is not a force in the ordinary sense of the word(see later). We shall, however, continue to use this unfortunate terminology, because it is customary and convenient.

We see that the term $l^2/2\mu r^2$ can be interpreted as the *centrifugal potential energy* of the particle and, as such, can be included with U(r) in an *effective potential energy* defined by

$$\boxed{V(r) = U(r) + \frac{l^2}{2\mu r^2}} \tag{8.34}$$

V(r) is therefore a *fictitious* potential that combines the real potential function U(r) with the energy term associated with the angular motion about the center of force. For the case of inverse-square-law central-force motion, the force is given by

$$F(r) = -\frac{k}{r^2} \tag{8.35}$$

from which

$$U(r) = -\int F(r)\,dr = -\frac{k}{r} \tag{8.36}$$

The effective potential function for gravitational attraction is therefore

$$V(r) = -\frac{k}{r} + \frac{l^2}{2\mu r^2} \tag{8.37}$$

This effective potential and its components are shown in Figure 8-5.

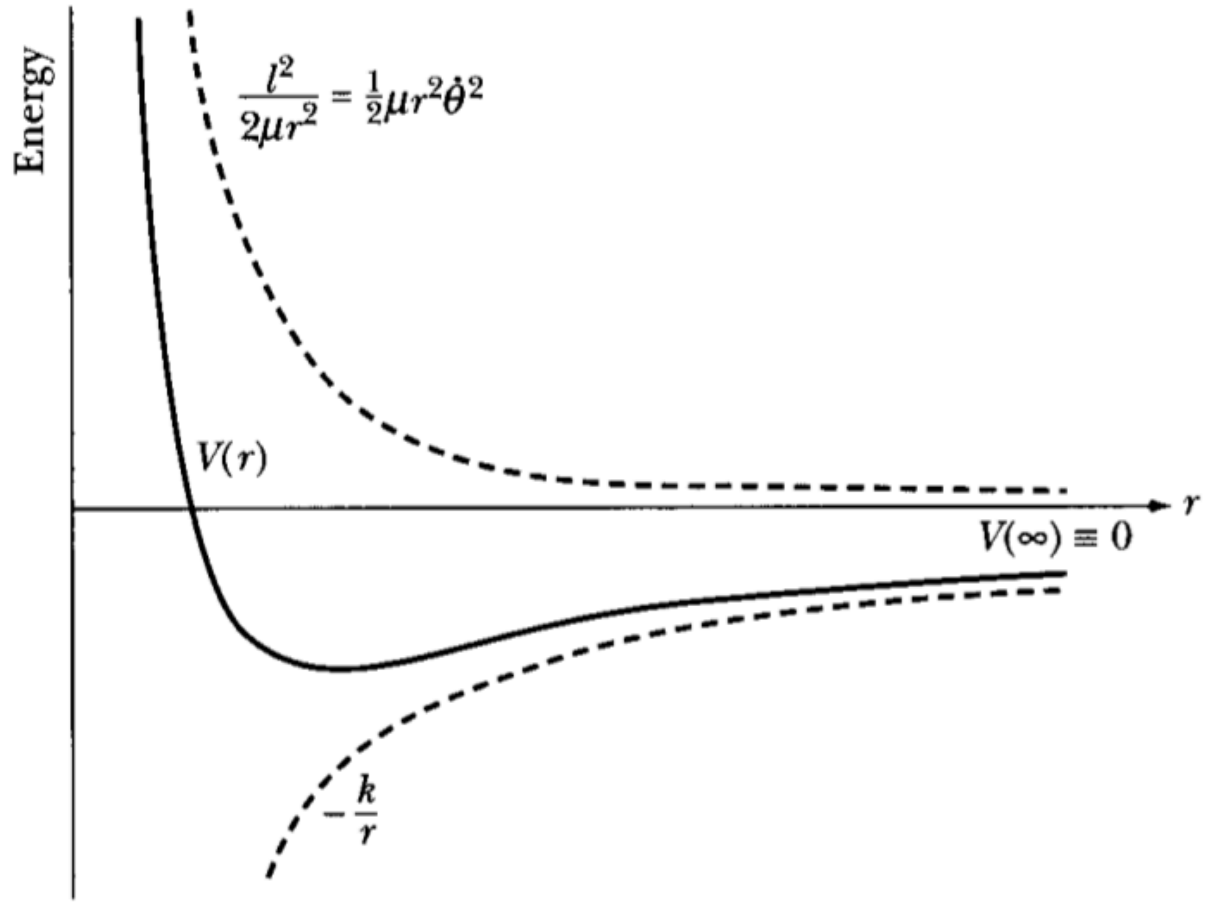


FIGURE 8-5 The effective potential for gravitational attraction $V(r)$ is composed of the real potential $-k/r$ term and the centrifugal potential energy $l^2/2\mu r^2$.

The value of the potential is arbitrarily taken to be zero at $r = \infty$. (This is implicit in Equation 8.36, where we omitted the constant of integration.)

We may now draw conclusions similar to those in Section 2.6 on the motion of a particle in an arbitrary potential well. If we plot the total energy E of the particle on a diagram similar to Figure 8-5, we may identify three regions of interest (see Figure 8-6).

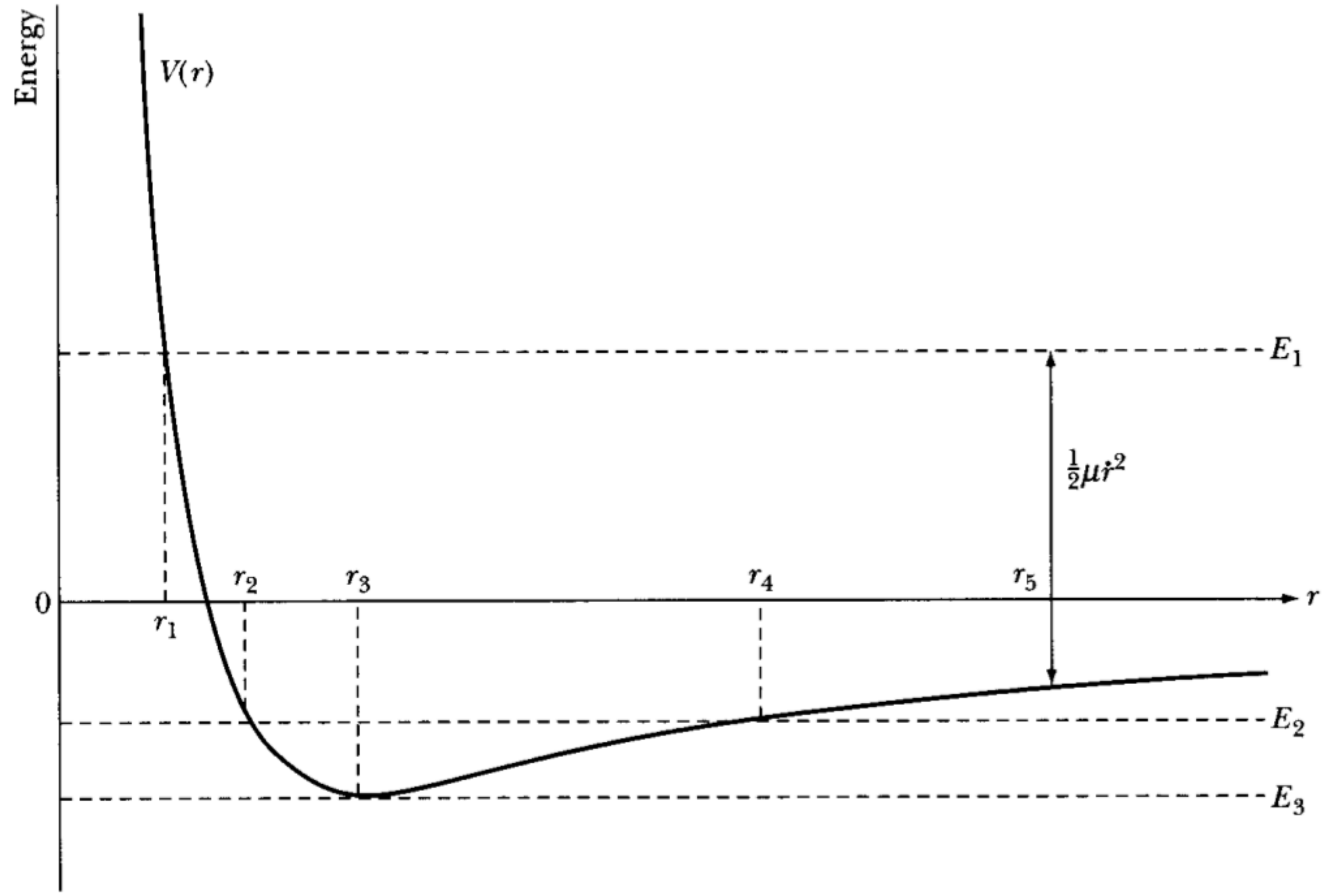


FIGURE 8-6 We can tell much about motion by looking at the total energy E on a potential energy plot. For example, for energy E_1 the particle's motion is unbounded. For energy E_2 the particle is bounded with $r_2 \leq r \leq r_4$. For energy E_3 the motion has $r = r_3$ and is circular.

If the total energy is positive or zero (e.g., $E_1 \geq 0$), then the motion is unbounded; the particle moves toward the force center (located at r = 0) from infinitely far away until it "strikes" the potential barrier at the turning point $r = r_1$ and is reflected back toward infinitely large r. Note that the height of the constant total energy line above V(r) at any r, such as r_5 in Figure 8-6, is equal to $\frac{1}{2}\mu\dot{r}^2$. Thus the radial velocity $\dot{r}$ vanishes and changes sign at the turning point (or points).

If the total energy is negative and lies between zero and the minimum value of V(r), as does E_2, then the motion is bounded, with $r_2 \leq r \leq r_4$.

The values r_2 and r_4 are the turning points, or the **apsidal distances**, of the orbit. If E equals the minimum value of the effective potential energy (see E_3 in Figure 8-6), then the radius of the particle's path is limited to the single value r_3, and then $\dot{r} = 0$ for all values of the time; hence the motion is circular.

Values of E less than $V_{min} = -(\mu k^2/2l^2)$ do not result in physically real motion; for such cases $\dot{r}^2 < 0$ and the velocity is imaginary.

The methods discussed in this section are often used in present-day research in general fields, especially atomic, molecular, and nuclear physics. For example, Figure 8-7 shows effective total nucleus-nucleus potentials for the scattering of ^{28}Si and ^{12}C.

The total potential includes the coulomb, nuclear, and the centrifugal contributions. The potential for $l = 0\hbar$ indicates the potential with no centrifugal term. For a relative angular momentum value of $l = 20\hbar$, a "pocket" exists where the two scattering nuclei may be bound together (even if only for a short time). For $l = 25\hbar$, the centrifugal "barrier" dominates, and the nuclei cannot form a bound state at all.

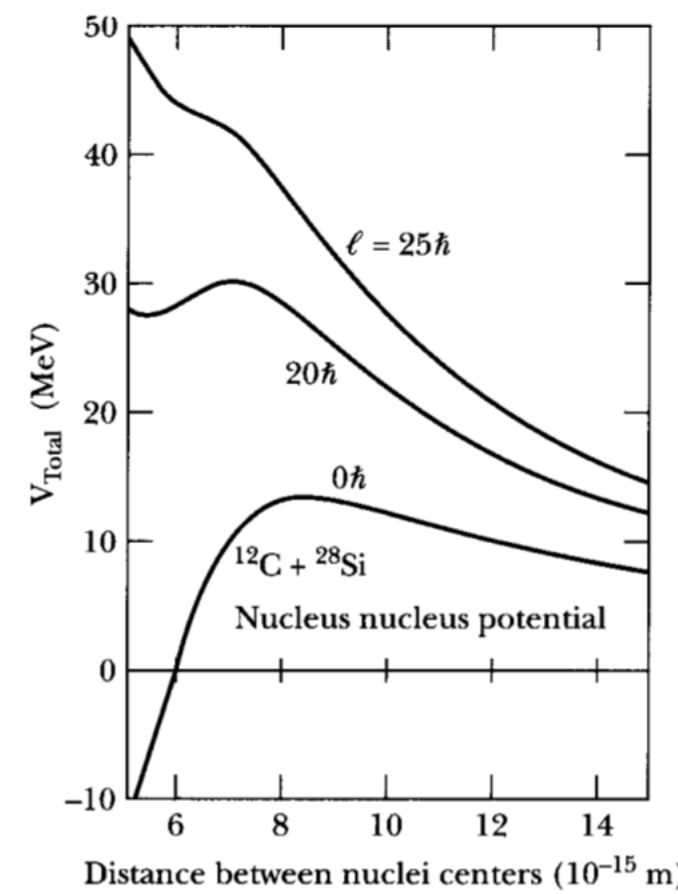


FIGURE 8-7 The total potential (coulomb, nuclear, and centrifugal) for scattering ^{28}Si nuclei from ^{12}C for various angular momentum l values as a function of distance between nuclei. For $l = 20\hbar$ a shallow pocket exists where the two nuclei may be bound together for a short time. For $l = 25\hbar$ the nuclei are not bound together.

8.7 Planetary Motion - Kepler's Problem

The equation for the path of a particle moving under the influence of a central force whose magnitude is inversely proportional to the square of the distance between the particle and the force center can be obtained (see Equation 8.17) from

$$\theta(r) = \int \frac{(l/r^2)\,dr}{\sqrt{2\mu\left(E + \frac{k}{r} - \frac{l^2}{2\mu r^2}\right)}} + \text{constant} \tag{8.38}$$

The integral can be evaluated if the variable is changed to $u = l/r$. If we define the origin of θ so that the minimum value of r is at $\theta = 0$, we find

$$\cos\theta = \frac{\frac{l^2}{\mu k}\cdot\frac{1}{r} - 1}{\sqrt{1 + \frac{2El^2}{\mu k^2}}} \tag{8.39}$$

Let us now define the following constants:

$$\left.\begin{aligned} \alpha &\equiv \frac{l^2}{\mu k} \\ \varepsilon &\equiv \sqrt{1 + \frac{2El^2}{\mu k^2}} \end{aligned}\right\} \tag{8.40}$$

Equation 8.39 can thus be written as

$$\boxed{\frac{\alpha}{r} = 1 + \varepsilon\cos\theta} \tag{8.41}$$

This is the equation of a conic section with one focus at the origin. The quantity ε is called the **eccentricity**, and 2α is termed the **latus rectum** of the orbit. Conic sections are formed by the intersection of a plane and a cone. A conic section is formed by the loci of points (formed in a plane), where the ratio of the distance from a fixed point (the focus) to a fixed line (called the directrix) is a constant. The directrix for the parabola is shown in Figure 8-8 by the vertical dashed line, drawn so that $r/r' = 1$.

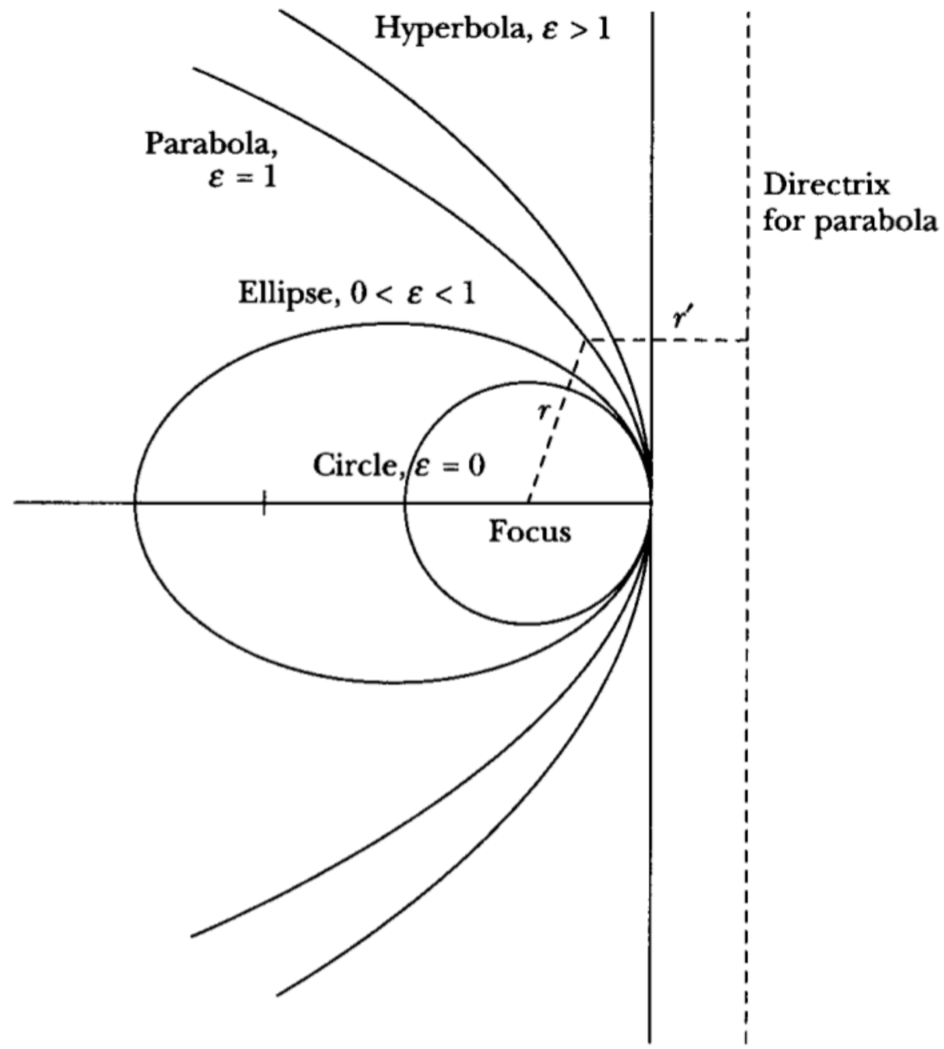


FIGURE 8-8 The orbits of the various conic sections are shown together with their eccentricities ε.

The minimum value for r in Equation 8.41 occurs when $\theta = 0$, or when $\cos\theta$ is a maximum. Thus the choice of the integration constant in Equation 8.38 corresponds to measuring θ from r_{min}, which position is called the **pericenter**; r_{max} corresponds to the **apocenter**. The general term for turning points is **apsides**. The corresponding terms for motion about the Sun are *perihelion* and *aphelion*, and for motion about Earth, *perigee* and *apogee*.

Various values of the eccentricity (and hence of the energy E) classify the orbits according to different conic sections (see Figure 8-8)

$\varepsilon > 1,$	$E > 0$	Hyperbola
$\varepsilon = 1,$	$E = 0$	Parabola
$0 < \varepsilon < 1,$	$V_{min} < E < 0$	Ellipse
$\varepsilon = 0,$	$E = V_{min}$	Circle

For planetary motion, the orbits are ellipses with major and minor axes (equal to 2a and 2b, respectively) given by

$$a = \frac{\alpha}{1-\varepsilon^2} = \frac{k}{2|E|} \tag{8.42}$$

$$b = \frac{\alpha}{\sqrt{1-\varepsilon^2}} = \frac{l}{\sqrt{2\mu|E|}} \tag{8.43}$$

Thus, the major axis depends only on the energy of the particle, whereas the minor axis is a function of both first integrals of the motion, E and L The geometry of elliptic orbits in terms of the parameters α, ε , a, and b is shown in Figure 8-9; P and P' are the foci.

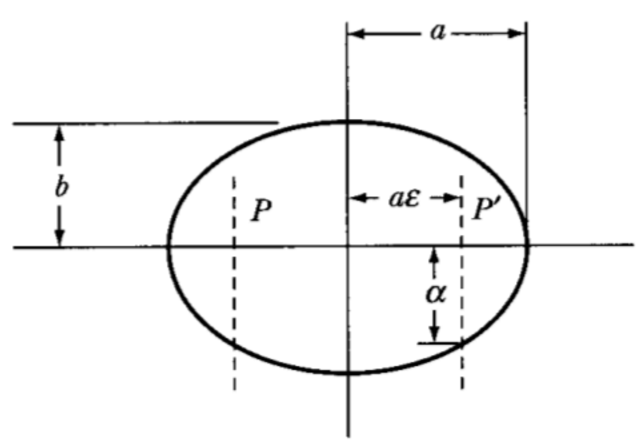


FIGURE 8-9 The geometry of elliptic orbits is shown in terms of parameters α, ε, a, and b. P and P' are the foci.

From this diagram, we see that the apsidal distances (r_{min} and r_{max} as measured from the foci to the orbit) are given by

$$\left.\begin{aligned} r_{\min} &= a(1-\varepsilon) = \frac{\alpha}{1+\varepsilon} \\ r_{\max} &= a(1+\varepsilon) = \frac{\alpha}{1-\varepsilon} \end{aligned}\right\} \tag{8.44}$$

To find the period for elliptic motion, we rewrite Equation 8.12 for the areal velocity as

$$dt = \frac{2\mu}{l}\,dA$$

Because the entire area A of the ellipse is swept out in one complete period τ ,

$$\begin{aligned} \int_0^\tau dt &= \frac{2\mu}{l}\int_0^A dA \\ \tau &= \frac{2\mu}{l}A \end{aligned} \tag{8.45}$$

The area of an ellipse is given by $A = \pi ab$, and using a and b from Equations 8.42 and 8.43, we find

$$\begin{aligned} \tau &= \frac{2\mu}{l}\cdot\pi ab = \frac{2\mu}{l}\cdot\pi\cdot\frac{k}{2|E|}\cdot\frac{l}{\sqrt{2\mu|E|}} \\ &= \pi k\sqrt{\frac{\mu}{2}}\cdot|E|^{-3/2} \end{aligned} \tag{8.46}$$

We also note from Equations 8.42 and 8.43 that the semiminor axis (the quantities a and b are called semimajor and semiminor axes, respectively.) can be written as

$$b = \sqrt{\alpha a} \tag{8.47}$$

Therefore, because $\alpha = l^2/\mu k$, the period τ can also be expressed as

$$\boxed{\tau^2 = \frac{4\pi^2\mu}{k} a^3} \tag{8.48}$$

This result, that the square of the period is proportional to the cube of the semimajor axis of the elliptic orbit, is known as **Kepler's Third Law**. Note that this result is concerned with the equivalent one-body problem, so account must be taken of the fact that it is the *reduced* mass μ that occurs in Equation 8.48. Kepler actually concluded that the squares of the periods of the planets were proportional to the cubes of the major axes of their orbits - with the same proportionality constant for all planets. In this sense, the statement is only approximately correct, because the reduced mass is different for each planet. In particular, because the gravitational force is given by

$$F(r) = -\frac{Gm_1 m_2}{r^2} = -\frac{k}{r^2}$$

we identify $k = Gm_1m_2$. The expression for the square of the period therefore becomes

$$\tau^2 = \frac{4\pi^2 a^3}{G(m_1 + m_2)} \cong \frac{4\pi^2 a^3}{Gm_2}, \qquad m_1 \ll m_2 \tag{8.49}$$

and Kepler's statement is correct only if the mass m_1 of a planet can be neglected with respect to the mass m_2 of the Sun. (But note, for example, that the mass of Jupiter is about 1/1000 of the mass of the Sun, so the departure from the approximate law is not difficult to observe in this case.)

Kepler's laws can now be summarized:

I. *Planets move in elliptical orbits about the Sun with the Sun at one focus.*

II. *The area per unit time swept out by a radius vector from the Sun to a planet is constant.*

III. *The square of a planet's period is proportional to the cube of the major axis of the planet's orbit.*

See Table 8-1 for some properties of the principal objects in the solar system.

TABLE 8-1 Some Properties of the Principal Objects in the Solar System

Name	Semimajor axis of orbit (in astronomical units[a])	Period (yr)	Eccentricity	Mass (in units of Earth's mass[b])
Sun	—	—	—	332,830
Mercury	0.3871	0.2408	0.2056	0.0552
Venus	0.7233	0.6152	0.0068	0.814
Earth	1.0000	1.0000	0.0167	1.000
Eros (asteroid)	1.4583	1.7610	0.2229	2×10^{-9} (?)
Mars	1.5237	1.8809	0.0934	0.1074
Ceres (asteroid)	*c*	4.6035	0.0789	1/8000 (?)
Jupiter	5.2028	*c*	0.0483	317.89
Saturn	9.5388	29.456	0.0560	*c*
Uranus	19.191	84.07	0.0461	14.56
Neptune	30.061	164.81	0.0100	17.15
Pluto	39.529	248.53	0.2484	0.002
Halley (comet)	18	76	0.967	$\sim 10^{-10}$

[a] One astronomical unit (A.U.) is the length of the semimajor axis of Earth's orbit. One A.U. $\cong 1.495 \times 10^{11}$ m $\cong 93 \times 10^{6}$ miles.

[b] Earth's mass is approximately 5.976×10^{24} kg.

EXAMPLE 8.4

Halley's comet, which passed around the sun early in 1986, moves in a highly elliptical orbit with an eccentricity of 0.967 and a period of 76 years. Calculate its minimum and maximum distances from the Sun.

Solution. Equation 8.49 relates the period of motion with the semimajor axes. Because m (Halley's comet) $\ll m_{sun}$,

$$a = \left(\frac{Gm_{\text{Sun}}\tau^2}{4\pi^2}\right)^{1/3}$$

$$= \left[\frac{\left(6.67 \times 10^{-11}\,\frac{\text{Nm}^2}{\text{kg}^2}\right)(1.99 \times 10^{30}\,\text{kg})\left(76\,\text{yr}\,\frac{365\,\text{day}}{\text{yr}}\,\frac{24\,\text{hr}}{\text{day}}\,\frac{3600\,\text{s}}{\text{hr}}\right)^2}{4\pi^2}\right]^{1/3}$$

$$a = 2.68 \times 10^{12}\ \text{m}$$

Using Equation 8.44, we can determine r_{min} and r_{max}.

$$r_{\text{min}} = 2.68 \times 10^{12}\ \text{m}(1 - 0.967) = 8.8 \times 10^{10}\ \text{m}$$
$$r_{\text{max}} = 2.68 \times 10^{12}\ \text{m}(1 + 0.967) = 5.27 \times 10^{12}\ \text{m}$$

This orbit takes the comet inside the path of Venus, almost to Mercury's orbit, and out past even the orbit of Neptune and sometimes even to the moderately eccentric orbit of Pluto. Edmond Halley is generally given the credit for bringing Newton's work on gravitational and central forces to the attention of the world. After observing the comet personally in 1682, Halley became interested. Partly as a result of a bet between Christopher Wren and Robert Hooke, Halley asked Newton in 1684 what paths the planets must follow if the Sun pulled them with a force inversely proportional to the square of their distances. To the astonishment of Halley, Newton replied, "Why, in ellipses, of course." Newton had worked it out 20 years previously but had not published the result. With painstaking effort, Halley was able in 1705 to predict the next occurrence of the comet, now bearing his name, to be in 1758.

8.8 Orbital Dynamics

The use of central-force motion is nowhere more useful, important, and interesting than in space dynamics. Although space dynamics is actually quite complex because of the gravitational attraction of a spacecraft to various bodies and the orbital motion involved, we examine two rather simple aspects: a proposed trip to Mars and flybys past comets and planets.

Orbits are changed by single or multiple thrusts of the rocket engines. The simplest maneuver is a single thrust applied in the orbital plane that does not change the direction of the angular momentum but does change the eccentricity and energy simultaneously. The most economical method of interplanetary transfer consists of moving from one circular heliocentric (Sun-oriented motion) orbit to another in the same plane. Earth and Mars represent such a system reasonably well, and a Hohmann transfer (Figure 8-10) represents the path of minimum total energy expenditure.

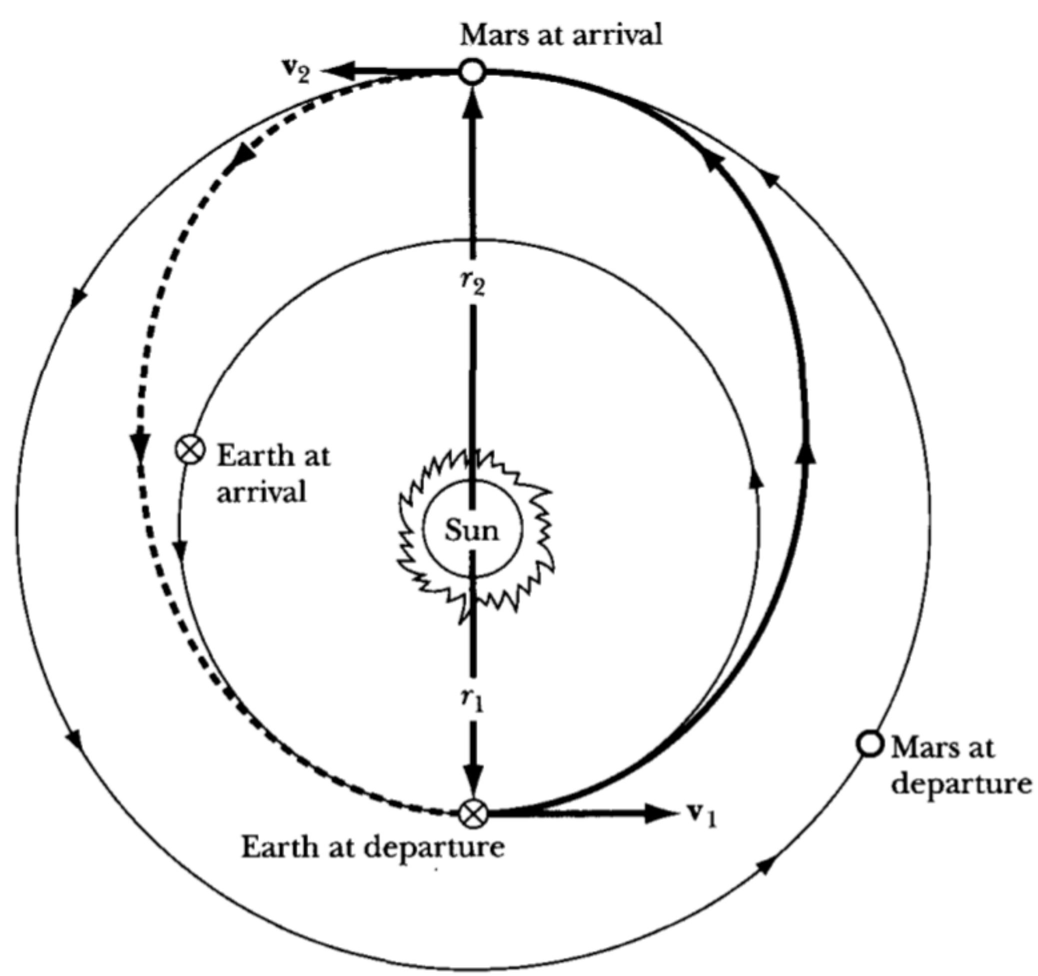


FIGURE 8-10 The Hohmann transfer for a round trip between Earth and Mars. It represents the minimum energy expenditure.

Two engine burns are required: A) the first burn injects the spacecraft from the circular Earth orbit to an elliptical transfer orbit that intersects Mars' orbit; B) the second burn transfers the spacecraft from the elliptical orbit into Mars' orbit.

We can calculate the velocity changes needed for a Hohmann transfer by calculating the velocity of a spacecraft moving in the orbit of Earth around the Sun (r_1 in Figure 8-10) and the velocity needed to "kick" it into an elliptical transfer orbit that can reach Mars' orbit. We are considering only the gravitational attraction of the Sun and not that of Earth and Mars.

For circles and ellipses we have, from Equation 8.42,

$$E = -\frac{k}{2a}$$

For a circular path around the Sun, this becomes

$$E = -\frac{k}{2r_1} = \frac{1}{2}mv_1^2 - \frac{k}{r_1} \tag{8.50}$$

where we have E = T+ U. We solve Equation 8.50 for v_1.

$$v_1 = \sqrt{\frac{k}{mr_1}} \tag{8.51}$$

We denote the semimajor axis of the transfer ellipse by a_t.

$$2a_t = r_1 + r_2$$

If we calculate the energy at the perihelion for the transfer ellipse, we have

$$E_t = \frac{-k}{r_1 + r_2} = \frac{1}{2}mv_{t1}^2 - \frac{k}{r_1} \tag{8.52}$$

where v_{t1} is the perihelion transfer speed. The direction of v_{t1} is along $\mathbf{v}_1$ in Figure 8-10. Solving Equation 8.52 for v_{t1} gives

$$v_{t1} = \sqrt{\frac{2k}{mr_1}\left(\frac{r_2}{r_1 + r_2}\right)} \tag{8.53}$$

The speed transfer Δv_1 needed is just

$$\Delta v_1 = v_{t1} - v_1 \tag{8.54}$$

Similarly, for the transfer from the ellipse to the circular orbit of radius r_2, we have

$$\Delta v_2 = v_2 - v_{t2} \tag{8.55}$$

where

$$v_2 = \sqrt{\frac{k}{mr_2}} \tag{8.56}$$

and

$$\left.\begin{aligned} v_{t2} &= \sqrt{\frac{2}{m}\left(E_t + \frac{k}{r_2}\right)} \\ v_{t2} &= \sqrt{\frac{2k}{mr_2}\left(\frac{r_1}{r_1 + r_2}\right)} \end{aligned}\right\} \tag{8.57}$$

The direction of v_{t2} is along $\mathbf{v}_2$ in Figure 8-10. The total speed increment can be determined by adding the speed changes, .

The total time required to make the transfer T_t is a half-period of the transfer orbit. From Equation 8.48, we have

$$\begin{aligned} T_t &= \frac{\tau_t}{2} \\ T_t &= \pi\sqrt{\frac{m}{k}}\, a_t^{3/2} \end{aligned} \tag{8.58}$$

EXAMPLE 8.5

Calculate the time needed for a spacecraft to make a Hohmann transfer from Earth to Mars and the heliocentric transfer speed required assuming both planets are in coplanar orbits

Solution. We need to insert the appropriate constants in Equation 8.58.

$$
\begin{aligned}
\frac{m}{k} &= \frac{m}{GmM_{\text{Sun}}} = \frac{1}{GM_{\text{Sun}}} \\
&= \frac{1}{(6.67 \times 10^{-11}\ \text{m}^3/\text{s}^2 \cdot \text{kg})(1.99 \times 10^{30}\ \text{kg})} \\
&= 7.53 \times 10^{-21}\ \text{s}^2/\text{m}^3
\end{aligned} \tag{8.59}
$$

Because k/m occurs so often in solar system calculations, we write it as well.

$$
\begin{aligned}
\frac{k}{m} &= 1.33 \times 10^{20}\ \text{m}^3/\text{s}^2 \\
a_t &= \frac{1}{2}(r_{\text{Earth}-\text{Sun}} + r_{\text{Mars}-\text{Sun}}) \\
&= \frac{1}{2}(1.50 \times 10^{11}\,\text{m} + 2.28 \times 10^{11}\,\text{m}) \\
&= 1.89 \times 10^{11}\,\text{m} \\
T_t &= \pi(7.53 \times 10^{-21}\ \text{s}^2/\text{m}^3)^{1/2}(1.89 \times 10^{11}\,\text{m})^{3/2} \\
&= 2.24 \times 10^{7}\,\text{s} \\
&= 259\ \text{days}
\end{aligned} \tag{8.60}
$$

The heliocentric speed needed for the transfer is given in Equation 8.53.

$$
\begin{aligned}
v_{t1} &= \left[\frac{2(1.33 \times 10^{20}\,\text{m}^3/\text{s}^2)(2.28 \times 10^{11}\,\text{m})}{(1.50 \times 10^{11}\,\text{m})(3.78 \times 10^{11}\,\text{m})}\right]^{1/2} \\
&= 3.27 \times 10^{4}\ \text{m/s} = 32.7\ \text{km/s}
\end{aligned}
$$

We can compare v_{t1} with the orbital speed of Earth (Equation 8.51).

$$v_1 = \left[\frac{1.33 \times 10^{20}\, \mathrm{m^3/s^2}}{1.50 \times 10^{11}\, \mathrm{m}}\right]^{1/2} = 29.8\ \mathrm{km/s}$$

For transfers to the outer planets, the spacecraft should be launched in the direction of Earth's orbit in order to gain Earth's orbital velocity. To transfer to the inner planets (e.g., to Venus), the spacecraft should be launched opposite Earth's motion. In each case, it is the relative velocity Δv_1 that is important to the spacecraft (i.e., relative to Earth).

Although the Hohmann transfer path represents the least energy expenditure, it does not represent the shortest time. For a round trip from Earth to Mars, the spacecraft would have to remain on Mars for 460 days until Earth and Mars were positioned correctly for the return trip (see Figure 8-1la).

The total trip B59 + 460 + 259 = 978 days = 2.7 yr) would probably be too long. Other schemes either use more fuel to gain speed (Figure 8-1lb) or use the slingshot effect of flybys. Such a flyby mission past Venus (see Figure 8-11c) could be done in less than 2 years with only a few weeks near (or on) Mars.

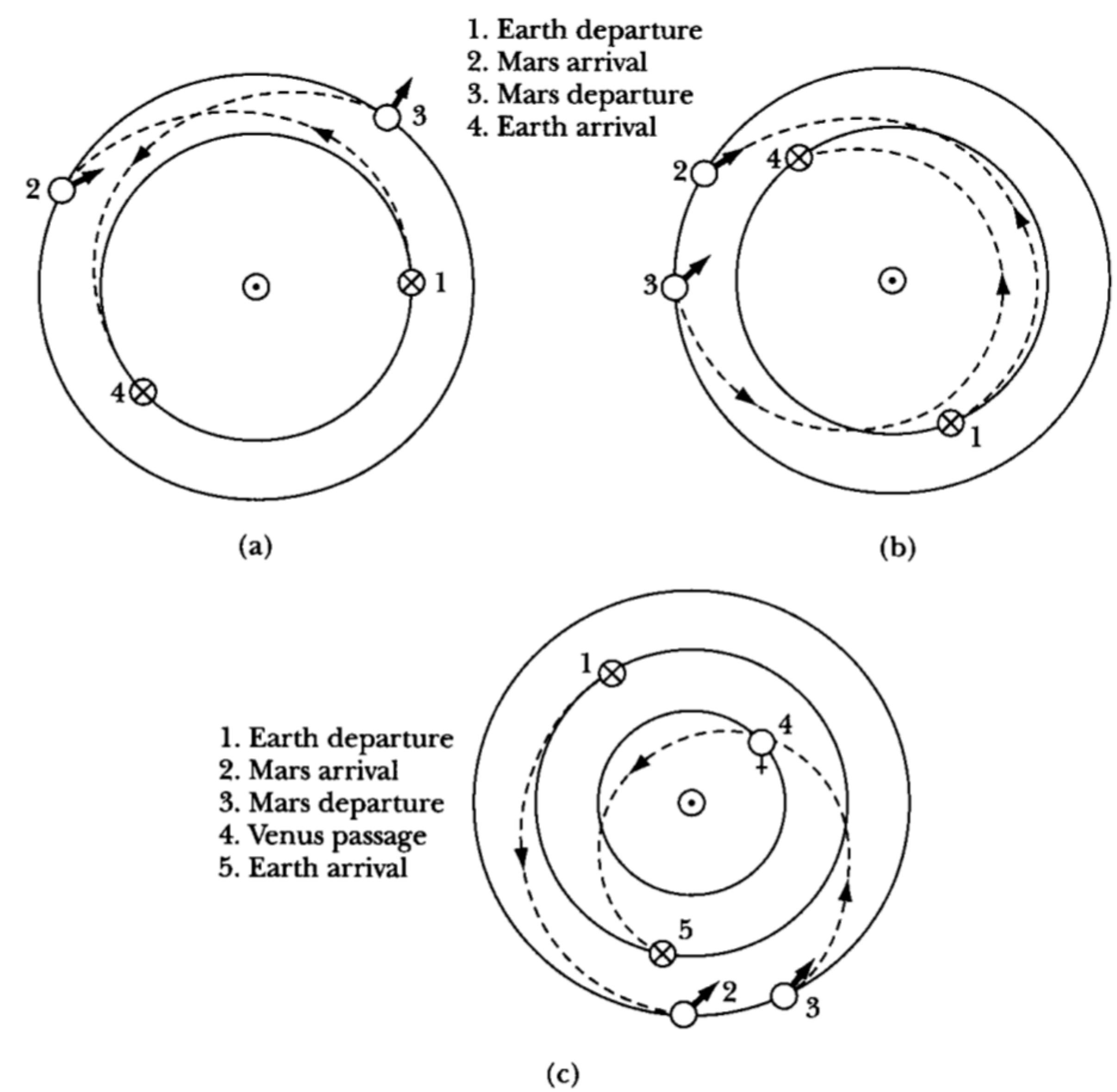


FIGURE 8-11 Round trips from Earth to Mars. (a) The minimum energy mission (Hohmann transfer) requires a long stopover on Mars before returning to Earth. (b) A shorter mission to Mars requires more fuel and a closer orbit to the Sun. (c) The fuel required for the shorter mission of (b) can be further improved if Venus is positioned for a gravity assist during flyby.

Several spacecraft in recent years have escaped Earth's gravitational attraction to explore our solar system. Such interplanetary transfer can be divided into three segments: A) the escape from Earth, B) a heliocentric transfer to the area of interest, and C) an encounter with another body - so far, either a planet or a comet. The spacecraft fuel required for such missions can be enormous, but a clever trick has been designed to "steal" energy from other solar system bodies. Because the mass of a spacecraft is so much smaller than the planets (or their moons), the energy loss of the heavenly body is negligible.

We examine a simple version of this flyby or slingshot effect that utilizes gravity assist. A spacecraft coming from infinity approaches a body (labeled B), interacts with B, and recedes. The path is a hyperbola (Figure 8-12). The initial and final velocities, *with respect to* B, are denoted by v_i' and v_f', respectively. The net effect on the spacecraft is a deflection angle of δ with respect to B.

If we examine the system in some inertial frame in which the motion of B occurs, the velocities of the spacecraft can be quite different *because of the motion of* B. The initial velocity v_i is shown in Figure 8-13a, and both v_i and v_f are shown in Figure 8-13b. Notice that the spacecraft has increased its speed as well as changed its direction. An increase in velocity occurs when the spacecraft passes behind B's direction of motion. Similarly, a decrease in velocity occurs when the spacecraft passes in *front* of B's motion.

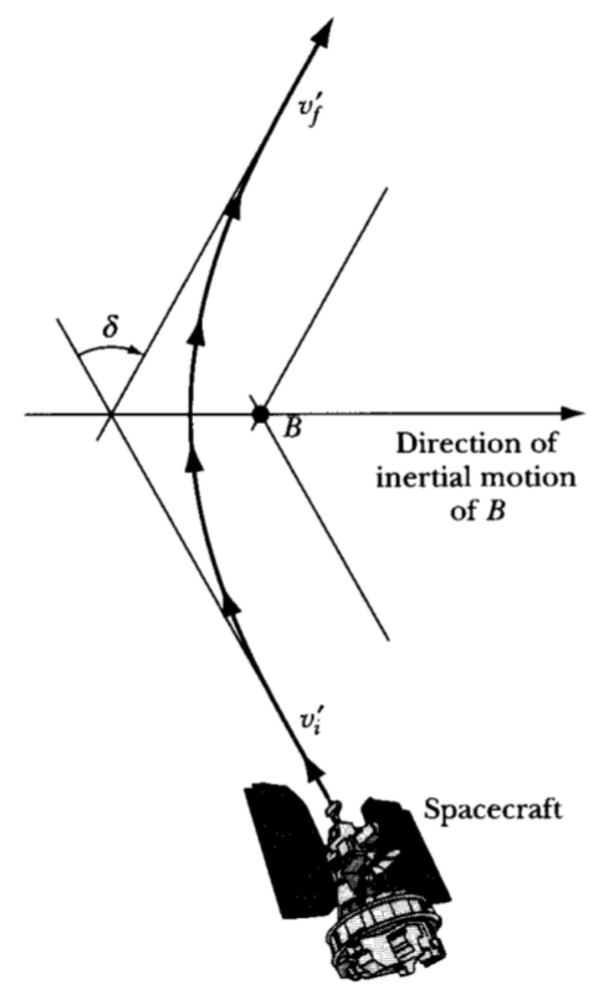


FIGURE 8-12 A spacecraft flies by a large body *B* (like a planet) and gains speed when it flies behind *B*'s direction of motion. Similarly, the spacecraft loses speed when it passes in front of *B*'s direction of motion. The direction of the spacecraft also changes.

During the 1970s, scientists at the Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA) realized that the four largest planets of our solar system would be in a fortuitous position to allow a spacecraft to fly past them and many of their 32 known moons in a single, relatively short "Grand Tour" mission using the gravity-assist method just discussed. This opportunity of the planets' alignment would not occur again for 175 years.

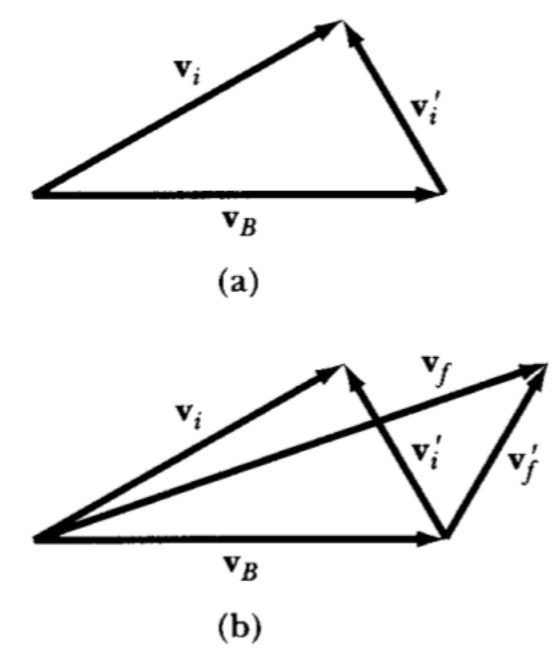


FIGURE 8-13 The vectors $\mathbf{v}_i'$ and $\mathbf{v}_f'$ are the initial and final velocities of the spacecraft with respect to B. The vectors $\mathbf{v}_i$ and $\mathbf{v}_f$ are the velocities in an inertial frame. (a) $\mathbf{v}_i = \mathbf{v}_B + \mathbf{v}_i'$. (b) $\mathbf{v}_f = \mathbf{v}_B + \mathbf{v}_f'$.

Because of budget constraints, there was not time to develop the new technology needed, and a mission to last only 4 years to visit just Jupiter and Saturn was approved and planned. No special equipment was put on board the twin *Voyager* spacecrafts for an encounter with Uranus and Neptune. *Voyagers 1 and 2* were launched in 1977 for visits to Jupiter in 1979 and Saturn in 1980 (*Voyager 1*) and 1981 (*Voyager 2*). Because of the success of these visits to Jupiter and Saturn, funding was later approved to extend *Voyager 2*'s mission to include Uranus and Neptune. The *Voyagers* are now on their way out of our solar system.

The path of *Voyager 2* is shown in Figure 8-14. The slingshot effect of gravity allowed the path of *Voyager 2* to be redirected, for example, toward Uranus as it passed Saturn by the method shown in Figure 8-12. The gravitational attraction from Saturn was used to pull the spacecraft off its straight path and redirect it at a different angle. The effect of the orbital motion of Saturn allows an increase in the spacecraft's speed. It was only by using this gravity-assist technique that the spectacular mission of *Voyager 2* was made possible in only a brief 12-year period.

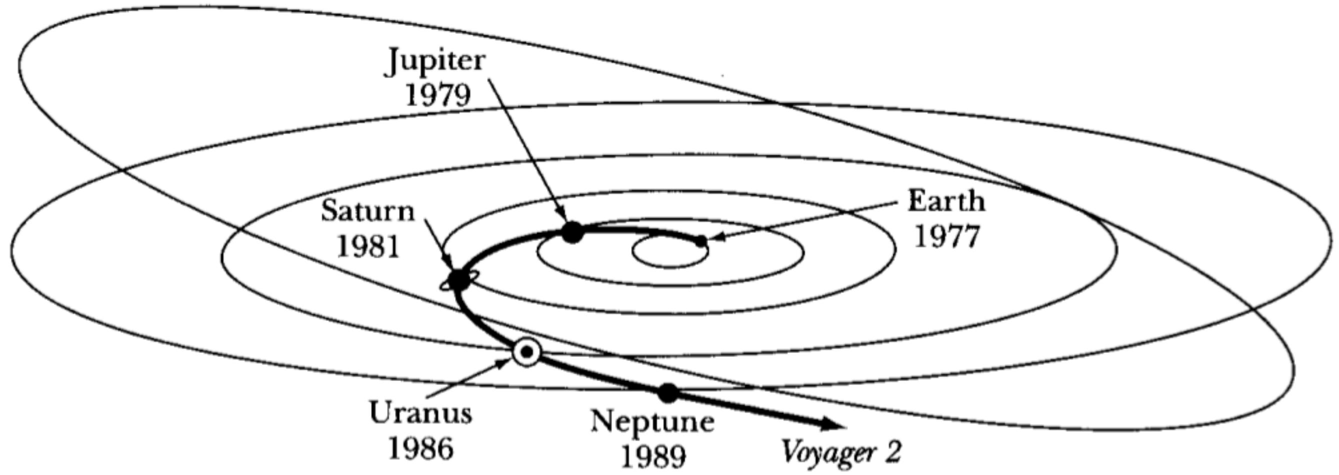


FIGURE 8-14 *Voyager 2* was launched in 1977 and passed by Jupiter, Saturn, Uranus, and Neptune. Gravitational assists were used in the mission.

Voyager 2 passed Uranus in 1986 and Neptune in 1989 before proceeding into interstellar space in one of the most successful space missions ever undertaken. Most planetary missions now take advantage of gravitational assists; for example, the Galileo satellite, which photographed the spectacular collisions of the Shoemaker-Levy comet with Jupiter in 1994 and reached Jupiter in 1995, was launched in 1989 but went by Earth twice (1990 and 1992) as well as Venus (1990) to gain speed and redirection.

A spectacular display of flybys occurred in the years 1982-1985 by a spacecraft initially called the International Sun-Earth Explorer 3 (ISEE-J). Launched in 1978, its mission was to monitor the solar wind between the Sun and Earth. For 4 years, the spacecraft circled in the elliptical plane about 2 million miles from Earth. In 1982 - because the United States had decided not to participate in a joint European, Japanese, and Soviet spacecraft investigation of Halley's comet in 1986 - NASA decided to reprogram the ISEE-3, renamed it the *International Cometary Explorer* (ICE), and sent it through the Giacobini-Zinner comet in September 1985, some 6 months before the flybys of other spacecraft with Halley's comet. The subsequent three-year journey of ICE was spectacular (Figure 8-15). The path of ICE included two close trips to Earth and five flybys of the moon along its billion-mile trip to the comet. During one flyby, the satellite came within 75 miles of the lunar surface. The entire path could be planned precisely because the force law is very well known. The eventual interaction with the comet, some 44 million miles from Earth, included a 20-minute trip through the comet - about 5,000 miles behind the comet's nucleus.

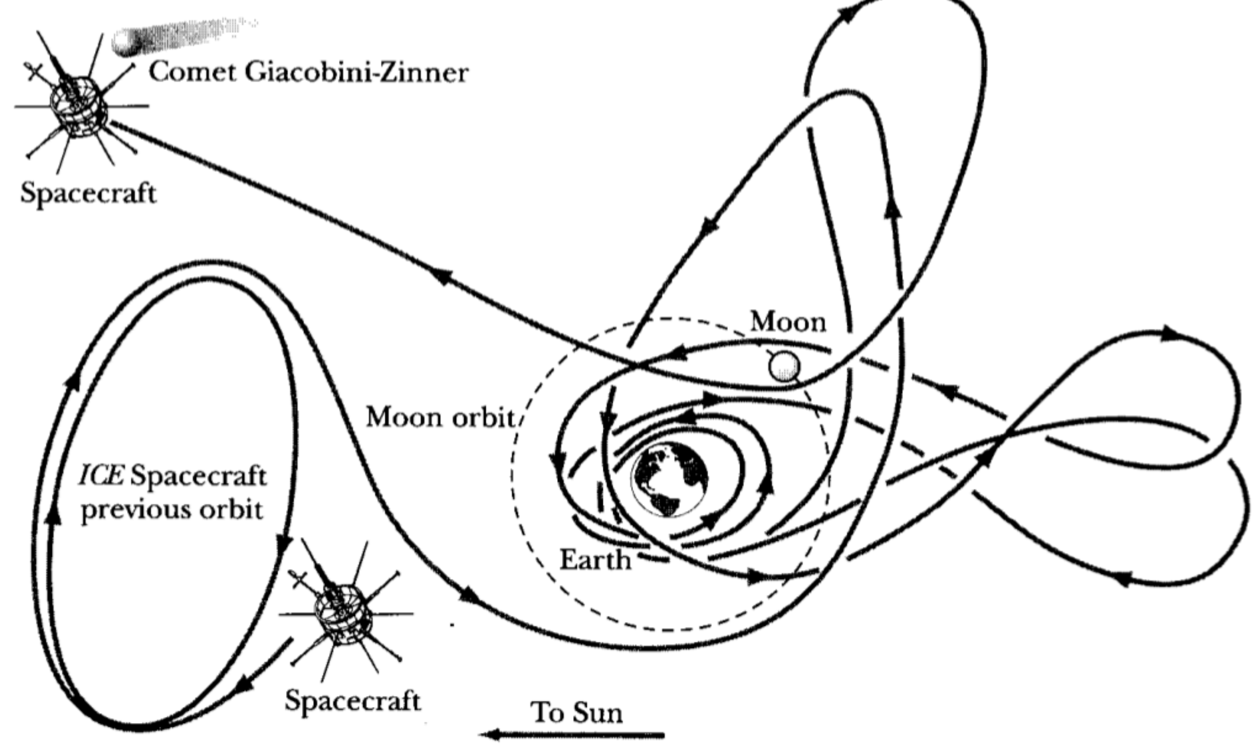


FIGURE 8-15 The NASA spacecraft initially called *ISEE-3* was reprogrammed to be the International Cometary Explorer and was sent on a spectacular three-year journey utilizing gravity assists on its way by the Comet Giacobini-Zinner.

8.9 Apsidal Angles and Precession

If a particle executes bounded, noncircular motion in a central-force field, then the radial distance from the force center to the particle must always be in the range $r_{max} \geq r \geq r_{min}$; that is, r must be bounded by the apsidal distances. Figure 8-6 indicates that only two apsidal distances exist for bounded, noncircular motion. But in executing one complete revolution in θ, the particle may not return to its original position (see Figure 8-4). The angular separation between two successive values of $r = r_{max}$ depends on the exact nature of the force. The angle between any two consecutive apsides is called the **apsidal angle**, and because a closed orbit must be symmetric about any apsis, it follows that all apsidal angles for such motion must be equal. The apsidal angle for elliptical motion, for example, is just π. If the orbit is not closed, the particle reaches the apsidal distances at different points in each revolution; the apsidal angle is not then a rational fraction of 2π, as is required for a closed orbit. If the orbit is *almost* closed, the apsides precess, or rotate slowly in the plane of the motion. This effect is exactly analogous to the slow rotation of the elliptical motion of a two-dimensional harmonic oscillator whose natural frequencies for the x and y motions are almost equal (see Section 3.3).

Because an inverse-square-law force requires that all elliptical orbits be exactly closed, the apsides must stay fixed in space for all time. If the apsides move with time, however slowly, this indicates that the force law under which the body moves does not vary exactly as the inverse square of the distance. This important fact was realized by Newton, who pointed out that any advance or regression of a planet's perihelion would require the radial dependence of the force law to be slightly different from $1/r^2$. Thus, Newton argued, the observation of the time dependence of the perihelia of the planets would be a sensitive test of the validity of the form of the universal gravitation law.

In point of fact, for planetary motion within the solar system, one expects that, because of the perturbations introduced by the existence of all the other planets, the force experienced by any planet does not vary exactly as $1/r^2$, if r is measured from the Sun. This effect is small, however, and only slight variations of planetary perihelia have been observed. The perihelion of Mercury, for example, which shows the largest effect, advances only about 574" of arc per century. This precession is in addition to the general precession of the equinox with respect to the "fixed" stars, which amounts to 5025.645" ± 0.050" per century Detailed calculations of the influence of the other planets on the motion of Mercury predict that the rate of advance of the perihelion should be approximately 531" per century. The uncertainties in this calculation are considerably less than the difference of 43" between observation and calculation, and for a considerable time, this discrepancy was the outstanding unresolved difficulty in the Newtonian theory. We now know that the modification introduced into the equation of motion of a planet by the general theory of relativity almost exactly accounts for the difference of 43". This result is one of the major triumphs of relativity theory.

We next indicate the way the advance of the perihelion can be calculated from the modified equation of motion. To perform this calculation, it is convenient to use the equation of motion in the form of Equation 8.20. If we use the universal gravitational law for F(r), we can write

$$\frac{d^2u}{d\theta^2} + u = -\frac{m}{l^2}\frac{1}{u^2}F(1/u)$$
$$= \frac{Gm^2M}{l^2} \tag{8.61}$$

where we consider the motion of a body of mass m in the gravitational field of a body of mass M. The quantity u is therefore the reciprocal of the distance between m and M.

The modification of the gravitational force law required by the general theory of relativity introduces into the force a small component that varies as $1/r^4 (= u^4)$. Thus, we have

$$\frac{d^2u}{d\theta^2} + u = \frac{Gm^2M}{l^2} + \frac{3GM}{c^2}u^2 \tag{8.62}$$

where c is the velocity of propagation of the gravitational interaction and is identified with the velocity of light. To simplify the notation, we define

$$\left.\begin{aligned} \frac{1}{\alpha} &\equiv \frac{Gm^2M}{l^2} \\ \delta &\equiv \frac{3GM}{c^2} \end{aligned}\right\} \tag{8.63}$$

and we can write Equation 8.62 as

$$\frac{d^2u}{d\theta^2} + u = \frac{1}{\alpha} + \delta u^2 \tag{8.64}$$

This is a nonlinear equation, and we use a successive approximation procedure to obtain a solution. We choose the first solution to be the solution of Equation 8.64 in the case that the term δu^2 is neglected (we eliminate the necessity of introducing an arbitrary phase into the argument of the cosine term by choosing to measure θ from the position of perihelion; i.e., u_1 is a maximum (and hence r_1 is a minimum) at $\theta = 0$:

$$u_1 = \frac{1}{\alpha}(1 + \varepsilon \cos \theta) \tag{8.65}$$

This is the familiar result for the pure inverse-square-law force (see Equation 8.41). Note that α is here the same as that defined in Equation 8.40 except that μ, has been replaced by m. If we substitute this expression into the right-hand side of Equation 8.64, we find

$$\begin{aligned} \frac{d^2u}{d\theta^2} + u &= \frac{1}{\alpha} + \frac{\delta}{\alpha^2}[1 + 2\varepsilon \cos \theta + \varepsilon^2 \cos^2 \theta] \\ &= \frac{1}{\alpha} + \frac{\delta}{\alpha^2}\left[1 + 2\varepsilon \cos \theta + \frac{\varepsilon^2}{2}(1 + \cos 2\theta)\right] \end{aligned} \tag{8.66}$$

where $\cos^2 \theta$ has been expanded in terms of $\cos 2\theta$. The first trial function u_1, when substituted into the left-hand side of Equation 8.64, reproduces only the first term on the right-hand side: $1/\alpha$. We can therefore construct a second trial function by adding to u_1 a term that reproduces the remainder of the right-hand side (in Equation 8.66). We can verify that such a particular integral is

$$u_p = \frac{\delta}{\alpha^2}\left[\left(1 + \frac{\varepsilon^2}{2}\right) + \varepsilon\theta \sin \theta - \frac{\varepsilon^2}{6} \cos 2\theta\right] \tag{8.67}$$

The second trial function is therefore $u_2 = u_1 + u_p$

If we stop the approximation procedure at this point, we have

$$\begin{aligned} u \cong u_2 &= u_1 + u_p \\ &= \left[\frac{1}{\alpha}(1 + \varepsilon \cos \theta) + \frac{\delta\varepsilon}{\alpha^2}\theta \sin \theta\right] \\ &\quad + \left[\frac{\delta}{\alpha^2}\left(1 + \frac{\varepsilon^2}{2}\right) - \frac{\delta\varepsilon^2}{6\alpha^2} \cos 2\theta\right] \end{aligned} \tag{8.68}$$

where we have regrouped the terms in u_1 and u_p.

Consider the terms in the second set of brackets in Equation 8.68: the first of these is just a constant, and the second is only a small and periodic disturbance of the normal Keplerian motion. Therefore, on a long time scale neither of these terms contributes, on the average, to any change in the positions of the apsides. But in the first set of brackets, the term proportional to θ produces secular and therefore observable effects. Let us consider the first set of brackets:

$$u_{\text{secular}} = \frac{1}{\alpha}\left[1 + \varepsilon\cos\theta + \frac{\delta\varepsilon}{\alpha}\theta\sin\theta\right] \tag{8.69}$$

Next, we can expand the quanuty

$$\begin{aligned} 1 + \varepsilon\cos\left(\theta - \frac{\delta}{\alpha}\theta\right) &= 1 + \varepsilon\left(\cos\theta\cos\frac{\delta}{\alpha}\theta + \sin\theta\sin\frac{\delta}{\alpha}\theta\right) \\ &\cong 1 + \varepsilon\cos\theta + \frac{\delta\varepsilon}{\alpha}\theta\sin\theta \end{aligned} \tag{8.70}$$

where we have used the fact that δ is small to approximate

$$\cos\frac{\delta}{\alpha}\theta \cong 1, \qquad \sin\frac{\delta}{\alpha}\theta \cong \frac{\delta}{\alpha}\theta$$

Hence, we can write u_{secular} as

$$u_{\text{secular}} \cong \frac{1}{\alpha}\left[1 + \varepsilon\cos\left(\theta - \frac{\delta}{\alpha}\theta\right)\right] \tag{8.71}$$

We have chosen to measure θ from the position of perihelion at $t = 0$. Successive appearances at perihelion result when the argument of the cosine
term in u_{secular} increases to 2π , 4π ,... , and so forth. But an increase of the argument by 2π requires that

$$\theta - \frac{\delta}{\alpha}\theta = 2\pi$$

or

$$\theta = \frac{2\pi}{1 - (\delta/\alpha)} \cong 2\pi\left(1 + \frac{\delta}{\alpha}\right)$$

Therefore, the effect of the relativistic term in the force law is to displace the perihelion in each revolution by an amount

$$\Delta \cong \frac{2\pi\delta}{\alpha} \tag{8.72a}$$

that is, the apsides rotate slowly in space. If we refer to the definitions of α and δ (Equations 8.63), we find

$$\Delta \cong 6\pi\left(\frac{GmM}{cl}\right)^2 \tag{8.72b}$$

TABLE 8-2 Precessional Rates for the Perihelia of Some Planets

Planet	Precessional rate (seconds of arc/century) Calculated	Observed
Mercury	43.03 ± 0.03	43.11 ± 0.45
Venus	8.63	8.4 ± 4.8
Earth	3.84	5.0 ± 1.2
Mars	1.35	—
Jupiter	0.06	—

From Equations 8.40 and 8.42, we can write $l^2 = \mu k a(1-\varepsilon^2)$: then, because k = GmM and $\mu \simeq m$, we have

$$\boxed{\Delta \cong \frac{6\pi GM}{ac^2(1 - \varepsilon^2)}}$$

We see therefore that the effect is enhanced if the semimajor axis a is small and if the eccentricity is large. Mercury, which is the planet nearest the sun and which has the most eccentric orbit of any planet (except Pluto), provides the most sensitive test of the theory. Alternatively, we can say that the relativistic advance of the perihelion is a maximum for Mercury because the orbital velocity is greatest for Mercury and the relativistic parameter v/c largest. The calculated value of the precessional rate for Mercury is 43.03" ± 0.03" of arc per century. The observed value (corrected for the influence of the other planets) is 43.11" ± 0.45", so the prediction of relativity theory is confirmed in striking fashion. The precessional rates for some of the planets are given in Table 8-2.

8.10 Stability of Circular Orbits

In Section 8.6, we pointed out that the orbit is circular if the total energy equals the minimum value of the effective potential energy, $E = V_{min}$. More generally, however, a circular orbit is allowed for any attractive potential, because the attractive force can always be made to just balance the centrifugal force by the proper choice of radial velocity. Although circular orbits are therefore always possible in a central, attractive force field, such orbits are not necessarily stable. A circular orbit at $r = \rho$ exists if $\dot{r}|_{r=\rho} = 0$ for all t; this is possible if $(\partial V/\partial r)|_{r=\rho} = 0$. But only if the effective potential has a true minimum does stability result. All other equilibrium circular orbits are unstable.

Let us consider an attractive central force with the form

$$F(r) = -\frac{k}{r^n} \tag{8.73}$$

The potential for such a force is

$$U(r) = -\frac{k}{n-1} \cdot \frac{1}{r^{(n-1)}} \tag{8.74}$$

and the effective potential function is

$$V(r) = -\frac{k}{n-1} \cdot \frac{1}{r^{(n-1)}} + \frac{l^2}{2\mu r^2} \tag{8.75}$$

The conditions for a minimum of V(r) and hence for a stable circular orbit with a radius ρ are

$$\left.\frac{\partial V}{\partial r}\right|_{r=\rho} = 0 \quad \text{and} \quad \left.\frac{\partial^2 V}{\partial r^2}\right|_{r=\rho} > 0 \tag{8.76}$$

Applying these criteria to the effective potential of Equation 8.75, we have

$$\left.\frac{\partial V}{\partial r}\right|_{r=\rho} = \frac{k}{\rho^n} - \frac{l^2}{\mu\rho^3} = 0$$

or

$$\rho^{(n-3)} = \frac{\mu k}{l^2} \tag{8.77}$$

and

$$\left.\frac{\partial^2 V}{\partial r^2}\right|_{r=\rho} = -\frac{nk}{\rho^{(n+1)}} + \frac{3l^2}{\mu\rho^4} > 0$$

so

$$-\frac{nk}{\rho^{(n-3)}} + \frac{3l^2}{\mu} > 0 \tag{8.78}$$

Substituting $\rho^{(n-3)}$ from Equation 8.77 into Equation 8.78, we have

$$(3 - n)\frac{l^2}{\mu} > 0 \tag{8.79}$$

The condition that a stable circular orbit exists is thus n < 3.

Next, we apply a more general procedure and inquire about the frequency of oscillation about a circular orbit in a general force field. We write the force as

$$F(r) = -\mu g(r) = -\frac{\partial U}{\partial r} \tag{8.80}$$

Equation 8.18 can now be written as

$$\ddot{r} - r\dot{\theta}^2 = -g(r) \tag{8.81}$$

Substituting for $\dot{\theta}$ from Equation 8.10,

$$\ddot{r} - \frac{l^2}{\mu^2 r^3} = -g(r) \tag{8.82}$$

We now consider the particle to be initially in a circular orbit with radius ρ and apply a perturbation of the form $r \to \rho + x$, where x is small. Because ρ = constant, we also have $\ddot{r} \to \ddot{\rho}$. Thus

$$\ddot{x} - \frac{l^2}{\mu^2 \rho^3 [1 + (x/\rho)]^3} = -g(\rho + x) \tag{8.83}$$

But by hypothesis $(x/\rho) \ll 1$, so we can expand the quantity:

$$[1 + (x/\rho)]^{-3} = 1 - 3(x/\rho) + \cdots \tag{8.84}$$

We also assume that g(r) = g(ρ + x) can be expanded in a Taylor series about the point r = ρ :

$$g(\rho + x) = g(\rho) + xg'(\rho) + \cdots \tag{8.85}$$

where

$$g'(\rho) \equiv \left.\frac{dg}{dr}\right|_{r=\rho}$$

If we neglect all terms in x^2 and higher powers, then the substitution of Equations 8.84 and 8.85 into Equation 8.83 yields

$$\ddot{x} - \frac{l^2}{\mu^2 \rho^3}[1 - 3(x/\rho)] \cong -[g(\rho) + xg'(\rho)] \tag{8.86}$$

Recall that we assumed the particle to be initially in a circular orbit with $r = \rho$. Under such a condition, no radial motion occurs - that is, $\dot{r}|_{r=\rho} = 0$. Then, also, $\ddot{r}|_{r=\rho} = 0$. Therefore, evaluating Equation 8.82 at $r = \rho$, we have

$$g(\rho) = \frac{l^2}{\mu^2 \rho^3} \tag{8.87}$$

Substituting this relation into Equation 8.86, we have, approximately,

$$\ddot{x} - g(\rho)[1 - 3(x/\rho)] \cong -[g(\rho) + xg'(\rho)]$$

or

$$\ddot{x} + \left[\frac{3g(\rho)}{\rho} + g'(\rho)\right]x \cong 0 \tag{8.88}$$

If we define

$$\omega_0^2 \equiv \frac{3g(\rho)}{\rho} + g'(\rho) \tag{8.89}$$

then Equation 8.88 becomes the familiar equation for the undamped harmonic oscillator:

$$\ddot{x} + \omega_0^2 x = 0 \tag{8.90}$$

The solution to this equation is

$$x(t) = Ae^{+i\omega_0 t} + Be^{-i\omega_0 t} \tag{8.91}$$

If $\omega_0^2 < 0$, so that ω_0 is imaginary, then the second term becomes $B\exp(|\omega_0|t)$, which clearly increases without limit as time increases. The condition for oscillation is therefore $\omega_0^2 > 0$, or

$$\frac{3g(\rho)}{\rho} + g'(\rho) > 0 \tag{8.92a}$$

Because $g(\rho) > 0$ (see Equation 8.87), we can divide through by $g(\rho)$ and write this inequality as

$$\frac{g'(\rho)}{g(\rho)} + \frac{3}{\rho} > 0 \tag{8.92b}$$

or, because g(r) and F(r) are related by a constant multiplicative factor, stability results if

$$\boxed{\frac{F'(\rho)}{F(\rho)} + \frac{3}{\rho} > 0} \tag{8.93}$$

We now compare the condition on the force law imposed by Equation 8.93 with that previously obtained for a power-law force:

$$F(r) = -\frac{k}{r^n} \tag{8.94}$$

Equation 8.93 becomes

$$\frac{nk\rho^{-(n+1)}}{-k\rho^{-n}} + \frac{3}{\rho} > 0$$

or

$$(3 - n) \cdot \frac{1}{\rho} > 0 \tag{8.95}$$

and we are led to the same condition as before - that is, $n < 3$. (We must note, however, that the case $n = 3$ needs further examination.)

EXAMPLE 8.6

Investigate the stability of circular orbits in a force field described by the potential function

$$U(r) = \frac{-k}{r} e^{-(r/a)} \tag{8.96}$$

where $k > 0$ and $a > 0$.

Solution. This potential is called the **screened Coulomb potential** (when $k = Ze^2/4\pi\varepsilon_0$, where Z is the atomic number and e is the electron charge) because it falls off with distance more rapidly than 1/r and hence approximates the electrostatic potential of the atomic nucleus in the vicinity of the nucleus by taking into account the partial "cancellation" or "screening" of the nuclear by the atomic electrons. The force is found from

$$F(r) = -\frac{\partial U}{\partial r} = -k\left(\frac{1}{ar} + \frac{1}{r^2}\right)e^{-(r/a)}$$

and

$$\frac{\partial F}{\partial r} = k\left(\frac{1}{a^2 r} + \frac{2}{ar^2} + \frac{2}{r^3}\right)e^{-(r/a)}$$

The condition for stability (see Equation 8.93) is

$$3 + \rho\frac{F'(\rho)}{F(\rho)} > 0$$

Therefore

$$3 + \frac{\rho k\left(\frac{1}{a^2\rho} + \frac{2}{a\rho^2} + \frac{2}{\rho^3}\right)}{-k\left(\frac{1}{a\rho} + \frac{1}{\rho^2}\right)} > 0$$

which simplifies to

$$a^2 + a\rho - \rho^2 > 0$$

We may write this as

$$\frac{a^2}{\rho^2} + \frac{a}{\rho} - 1 > 0$$

Stability thus results for all $q \equiv a/\rho$ that exceed the value satisfying the equation

$$q^2 + q - 1 = 0$$

The positive (and therefore the only physically meaningful) solution is

$$q = \frac{1}{2}(\sqrt{5} - 1) \cong 0.62$$

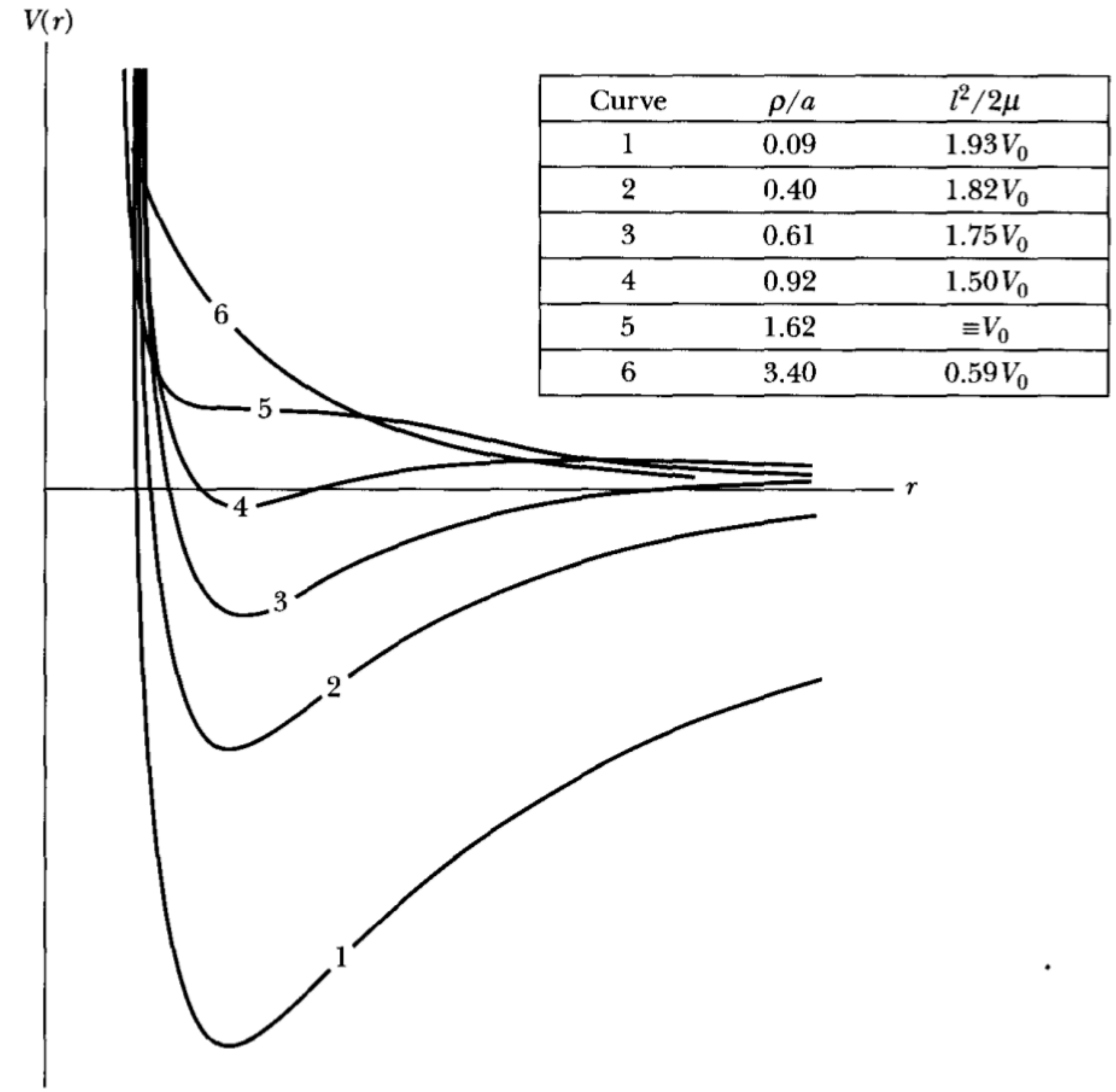
If, then, the angular momentum and energy allow a circular orbit at $r = \rho$, the motion is stable if

$$\frac{a}{\rho} \gtrsim 0.62$$

or

$$\rho \lesssim 1.62a \tag{8.97}$$

The stability condition for orbits in a screened potential is illustrated graphically in Figure 8-16, which shows the potential V(r) for various values of ρ/a. The force constant k is the same for all the curves, but $l^2/2\mu$ has been adjusted to maintain the minimum of the potential at the same value of the radius as a is changed.



Curve	ρ/a	$l^2/2\mu$
1	0.09	$1.93V_0$
2	0.40	$1.82V_0$
3	0.61	$1.75V_0$
4	0.92	$1.50V_0$
5	1.62	$\equiv V_0$
6	3.40	$0.59V_0$

FIGURE 8-16 Example 8.7. Potentials 1–4 produce a stable, circular orbit for values of $\rho/a \lesssim 1.62$.

For $\rho/a < 1.62$, a true minimum exists for the potential, indicating that the circular orbit is stable with respect to small oscillations. For $\rho/a > 1.62$, there is no minimum, so circular orbits cannot exist. For $\rho/a = 1.62$, the potential has zero slope at the position that a circular orbit would occupy. The orbit is unstable at this position, because (ω_0^2 is zero in Equation 8.90 and the displacement x increases linearly with time.

An interesting feature of this potential function is that under certain conditions there can exist bound orbits for which the total energy is positive (see, for example, curve 4 in Figure 8-16).

EXAMPLE 8.7

Determine whether a particle moving on the inside surface of a cone under the influence of gravity (see Example 7.4) can have a stable circular orbit.

Solution. In Example 7.4, we found that the angular momentum about the z-axis was a constant of the motion:

$$l = mr^2\dot{\theta} = \text{constant}$$

We also found the equation of motion for the coordinate r:

$$\ddot{r} - r\dot{\theta}^2 \sin^2\alpha + g \sin\alpha \cos\alpha = 0 \tag{8.98}$$

If the initial conditions are appropriately selected, the particle can move in a circular orbit about the vertical axis with the plane of the orbit at a constant height z_0 above the horizontal plane passing through the apex of the cone. Although this problem does not involve a central force, certain aspects of the motion are the same as for the central-force case. Thus we may discuss, for example, the stability of circular orbits for the particle. To do this, we perform a perturbation calculation.

First, we assume that a circular orbit exists for $r = \rho$. Then, we apply the perturbation $r \rightarrow \rho + x$. The quantity $r\dot{\theta}^2$ in Equation 8.98 can be expressed as

$$
\begin{aligned}
r\dot{\theta}^2 &= r\cdot\frac{l^2}{m^2r^4} = \frac{l^2}{m^2r^3} \\
&= \frac{l^2}{m^2}(\rho + x)^{-3} = \frac{l^2}{m^2\rho^3}\left(1 + \frac{x}{\rho}\right)^{-3} \\
&\cong \frac{l^2}{m^2\rho^3}\left(1 - 3\frac{x}{\rho}\right)
\end{aligned}
$$

where we have retained only the first term in the expansion, because x/ρ is by hypothesis a small quantity.

Then, because $\ddot{\rho} = 0$, Equation 8.98 becomes, approximately,

$$\ddot{x} - \frac{l^2 \sin^2 \alpha}{m^2\rho^3}\left(1 - 3\frac{x}{\rho}\right) + g \sin \alpha \cos \alpha = 0$$

or

$$\ddot{x} + \left(\frac{3l^2 \sin^2 \alpha}{m^2\rho^4}\right)x - \frac{l^2 \sin \alpha}{m^2\rho^3} + g \sin \alpha \cos \alpha = 0 \tag{8.99}$$

If we evaluate Equation 8.98 at $r = \rho$, then $\ddot{r} = 0$, and we have

$$
\begin{aligned}
g \sin \alpha \cos \alpha &= \rho\dot{\theta}^2 \sin^2 \alpha \\
&= \frac{l^2}{m^2\rho^3} \sin^2 \alpha
\end{aligned}
$$

In view of this result, the last two terms in Equation 8.99 cancel, and there remains

$$\ddot{x} + \left(\frac{3l^2 \sin^2 \alpha}{m^2\rho^4}\right)x = 0$$

The solution to this equation is just a harmonic oscillation with a frequency ω where

$$\omega = \frac{\sqrt{3}l}{m\rho^2} \sin \alpha$$

Thus, the circular orbit is stable.

CHAPTER 14 - Special Theory of Relativity

14.1 Introduction

In Section 2.7, it was pointed out that the Newtonian idea of the complete separability of space and time and the concept of the absoluteness of time break down when they are subjected to critical analysis. The final overthrow of the Newtonian system as the ultimate description of dynamics was the result of several crucial experiments, culminating with the work of Michelson and Morley in 1881-1887. The results of these experiments indicated that the speed of light is independent of any relative uniform motion between source and observer. This fact, coupled with the finite speed of light, required a fundamental reorganization of the structure of dynamics. This was provided during the period 1904-1905 by H. Poincare, H. A. Lorentz, and A. Einstein, who formulated the theory of relativity in order to provide a consistent description of the experimental facts. The basis of relativity theory is contained in two postulates:

I. *The laws of physical phenomena are the same in all inertial reference frames (that is, only the relative motion of inertial frames can be measured; the concept of motion relative to "absolute rest" is meaningless).*

II. *The velocity of light (in free space) is a universal constant, independent of any relative motion of the source and the observer.*

Using these postulates as a foundation, Einstein was able to construct a beautiful, logically precise theory. A wide variety of phenomena that take place at high velocity and cannot be interpreted in the Newtonian scheme are accurately described by relativity theory.

Postulate I, which Einstein called the *principle of relativity*, is the fundamental basis for the theory of relativity. Postulate II, the law of propagation of light, follows from Postulate I if we accept, as Einstein did, that Maxwell's equations are fundamental laws of physics. Maxwell's equations predict the speed of light in vacuum to be c, and Einstein believed this to be the case in all inertial reference frames.

We do not attempt here to give the experimental background for the theory of relativity; such information can be found in essentially every textbook on modem physics and in many others concerned with electrodynamics. Rather, we simply accept as correct the above two postulates and work out some of their consequences for the area of mechanics. The discussion here is limited to the case of **special relativity**, in which we consider only inertial reference frames, that is, frames that are in uniform motion with respect to one another. The more general treatment of accelerated reference frames is the subject of the **general theory of relativity**.

14.2 Galilean Invariance

In Newtonian mechanics, the concepts of space and time are completely separable; furthermore, time is assumed to be an absolute quantity susceptible of precise definition independent of the reference frame. These assumptions lead to the invariance of the laws of mechanics under coordinate transformations of the following type. Consider two inertial reference frames K and K', which move along their x_1 and x'_1-axes with a uniform relative velocity v (Figure 14-1).

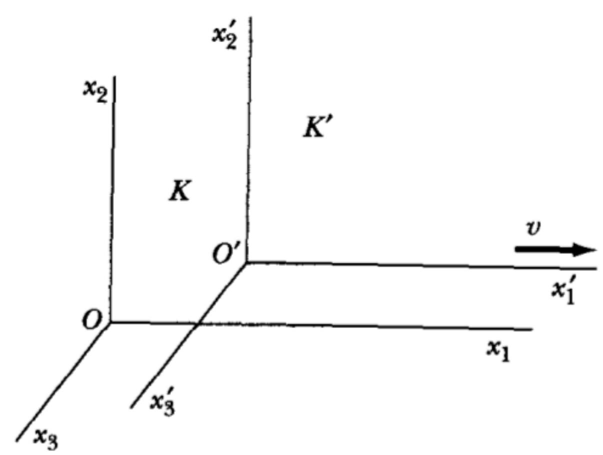


FIGURE 14-1 Two inertial reference frames K and K' move along their x_1- and x'_1-axes with a uniform relative velocity v.

The transformation of the coordinates of a point from one system to the other is clearly of the form

$$\left.\begin{aligned} x'_1 &= x_1 - vt \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned}\right\} \tag{14.1a}$$

Also, we have

$$t' = t \tag{14.1b}$$

Equations 14.1 define a **Galilean transformation**. Furthermore, the element of length in the two systems is the same and is given by

$$\begin{aligned} ds^2 &= \sum_j dx_j^2 \\ &= \sum_j dx_j'^2 = ds'^2 \end{aligned} \tag{14.2}$$

The fact that Newton's laws are invariant with respect to Galilean transformations is termed **the principle of Newtonian relativity** or **Galilean invariance**. Newton's equations of motion in the two systems are

$$\begin{aligned} F_j &= m\ddot{x}_j \\ &= m\ddot{x}_j' = F_j' \end{aligned} \tag{14.3}$$

The form of the law of motion is then *invariant* to a Galilean transformation. The individual terms are not invariant, however, but they transform according to the same scheme and are said to be *covariant*.

We can easily show that the Galilean transformation is inconsistent with Postulate II. Consider a light pulse emanating from a flashbulb positioned in frame K'. The velocity transformation is found from Equation 14.1a, where we consider the light pulse only along x_1:

$$\dot{x}_1' = \dot{x}_1 - v \tag{14.4}$$

In system K', the velocity is measured $\dot{x}_1' = c$; Equation 14.4 therefore indicates the speed of the light pulse to be $\dot{x}_1 = c + v$, clearly in violation of Postulate II.

14.3 Lorentz Transformation

The principle of Galilean invariance predicts that the velocity of light is different in two inertial reference frames that are in relative motion. This result is in contradiction to the second postulate of relativity. Therefore, a new transformation law that renders physical laws *relativistically* covariant must be found. Such a transformation law is the **Lorentz transformation**. The original use of the Lorentz transformation preceded the development of Einsteinian relativity theory, but it also follows from the basic postulates of relativity; we derive it on this basis in the following discussion.

If a light pulse from a flashbulb is emitted from the common origin of the systems K and K' (see Figure 14-1) when they are coincident, then according to Postulate II, the wavefronts observed in the two systems must be described''' by

$$\left.\begin{aligned} \sum_{j=1}^{3} x_j^2 - c^2t^2 = 0 \\ \sum_{j=1}^{3} x_j'^2 - c^2t'^2 = 0 \end{aligned}\right\} \tag{14.5}$$

We can already see that Equations 14.5, which are consistent with the two postulates of the theory of relativity, cannot be reconciled with the Galilean transformations of Equations 14.1. The Galilean transformation allows a spherical light wavefront in one system but requires the center of the spherical wavefront in the second system to move at velocity v with respect to the first system. The interpretation of Equations 14.5, according to Postulate II, is that each observer believes that his spherical wavefront has its center fixed at his own coordinate origin as the wavefront expands.

We are faced with a quandary. We must abandon either the two relativity postulates or the Galilean transformation. Much experimental evidence, including the Michelson-Morley experiment and the aberration of starlight, requires the two postulates. However, the belief in the Galilean transformation is entrenched in our minds by our everyday experience.

The Galilean transformation had produced satisfactory results, including those of the preceding chapters of this book, for centuries. Einstein's great contribution was to realize that the Galilean transformation was *approximately* correct, but that we needed to reexamine our concepts of space and time.

Notice that we do not assume t = t' in Equations 14.5. Each system, K and K', has its own clocks, and we assume that a clock may be located at any point in space. These clocks are all identical, run the same way, and are synchronized. Because the flashbulb goes off when the origins are coincident and the systems move only in the x_1-direction with respect to each other, by direct observation we have

$$\left.\begin{aligned} x_2' &= x_2 \\ x_3' &= x_3 \end{aligned}\right\} \tag{14.6}$$

At time t = t' = 0, when the flashbulb goes off, the motion of the origin O' of K' is measured in K to be

$$x_1 - vt = 0 \tag{14.7}$$

and in system K', the motion of O' is

$$x_1' = 0 \tag{14.8}$$

At time t = t' = 0 we have $x'_1 = x_1 - vt$, but we know that Equation 14.1a is incorrect. Let us assume the next simplest transformation, namely,

$$x_1' = \gamma(x_1 - vt) \tag{14.9}$$

where γ is some constant that may depend on v and some constants, but not on the coordinates x_1, x'_1, t, or t'.

Equation 14.9 is a linear equation and assures us that each event in K corresponds to one and only one event in K'. This additional assumption in our derivation will be vindicated if we can produce a transformation that is consistent with all the experimental results.

Notice that γ must normally be very close to 1 to be consistent with the classical results discussed in earlier chapters.

We can use the preceding arguments to describe the motion of the origin O of system K in both K and K' to also determine

$$x_1 = \gamma'(x_1' + vt') \tag{14.10}$$

where we only have to change the relative velocities of the two systems.

Postulate I demands that the laws of physics be the same in both reference systems such that $\gamma = \gamma'$. By substituting x'_1 from Equation 14.9 into Equation 14.10, we can solve the remaining equation for t':

$$t' = \gamma t + \frac{x_1}{\gamma v}(1 - \gamma^2) \tag{14.11}$$

Postulate II demands that the speed of light be measured to be the same in both systems. Therefore, in both systems we have similar equations for the position of the flashbulb light pulse:

$$\left.\begin{aligned} x_1 &= ct \\ x_1' &= ct' \end{aligned}\right\} \tag{14.12}$$

Algebraic manipulation of Equations 14.9-14.12 gives

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{14.13}$$

The complete transformation equations can now be written as

$$\left.\begin{aligned} x_1' &= \frac{x_1 - vt}{\sqrt{1 - v^2/c^2}} \\ x_2' &= x_2 \\ x_3' &= x_3 \\ t' &= \frac{t - \dfrac{vx_1}{c^2}}{\sqrt{1 - v^2/c^2}} \end{aligned}\right\} \tag{14.14}$$

These equations are known as the Lorentz (or Lorentz-Einstein) transformation in honor of the Dutch physicist H. A. Lorentz, who first showed that the equations are necessary so that the laws of electromagnetism have the same form in all inertial reference frames. Einstein showed that these equations are required for all the laws of physics.

The inverse transformation can easily be obtained by replacing v by -v and exchanging primed and unprimed quantities in Equations 14.14.

$$\left.\begin{aligned} x_1 &= \frac{x_1' + vt'}{\sqrt{1 - v^2/c^2}} \\ x_2 &= x_2' \\ x_3 &= x_3' \\ t &= \frac{t' + \dfrac{vx_1'}{c^2}}{\sqrt{1 - v^2/c^2}} \end{aligned}\right\} \tag{14.15}$$

As required, these equations reduce to the Galilean equations (Equations 14.1) when $v \to 0$ (or when $c \to \infty$).

In electrodynamics, the fields propagate with the speed of light, so Galilean transformations are never allowed. Indeed, the fact that the electrodynamic field equations (**Maxwell's equations**) are not covariant to Galilean transformations was a main factor in the realization of the need for a new theory. It seems rather extraordinary that Maxwell's equations, which are a complete set of equations for the electromagnetic field and are *covariant to Lorentz transformations*, were deduced from experiment long before the advent of relativity theory.

The velocities measured in each of the systems are denoted by u

$$\left.\begin{aligned} u_i &= \frac{dx_i}{dt} \\ u'_i &= \frac{dx'_i}{dt'} \end{aligned}\right\} \tag{14.16}$$

Using Equations 14.14, we determine

$$u'_1 = \frac{dx'_1}{dt'} = \frac{dx_1 - v\,dt}{dt - \dfrac{v}{c^2}dx_1}$$

$$u'_1 = \frac{u_1 - v}{1 - \dfrac{u_1 v}{c^2}} \tag{14.17a}$$

Similarly, we determine

$$u'_2 = \frac{u_2}{\gamma\left(1 - \dfrac{u_1 v}{c^2}\right)} \tag{14.17b}$$

$$u'_3 = \frac{u_3}{\gamma\left(1 - \dfrac{u_1 v}{c^2}\right)} \tag{14.17c}$$

Now we can determine whether Postulate II is satisfied directly. An observer in system K measures the speed of the light pulse from the flashbulb to be $u_1 = c$ in the x_1-direction. From Equation 14.17a, an observer in K' measures

$$u'_1 = \frac{c - v}{1 - \dfrac{v}{c}} = c\left(\frac{c - v}{c - v}\right) = c$$

as required by Postulate II, independent of the relative system speed v.

EXAMPLE 14.1

Determine the relativistic length contraction using the Lorentz transformation.

Solution. Consider a rod of length lying along the x_1-axis of an inertial frame K. An observer in system K' moving with uniform speed v along the x_1-axis (as in Figure 14-1) measures the length of the rod in the observer's own coordinate system by determining at a given instant of time t' the difference in the coordinates of the ends of the rod, $x'_1(2) - x'_1(1)$. According to the transformation equations (Equations 14.14),

$$x'_1(2) - x'_1(1) = \frac{[x_1(2) - x_1(1)] - v[t(2) - t(1)]}{\sqrt{1 - v^2/c^2}} \tag{14.18}$$

where $x_1(2) - x_1(1) = l$. Note that times t(2) and t(1) are the times in the K system at which the observations are made; they do not correspond to the instants in K' at which the observer measures the rod. In fact, because $t'(2) = t'(1)$, Equations 14.14 give

$$t(2) - t(1) = [x_1(2) - x_1(1)]\frac{v}{c^2}$$

The length l' as measured in the K' system is therefore

$$l' = x'_1(2) - x'_1(1)$$

Equation 14.18 now becomes

$$\text{length contraction} \quad \boxed{l' = l\sqrt{1 - v^2/c^2}} \tag{14.19}$$

and, to a stationary observer in K, objects in K' also appear contracted. Thus, to an observer in motion relative to an object, the dimensions of objects are contracted by a factor $\sqrt{1-\beta^2}$ in the direction of motion, in which $\beta = v/c$.

An interesting consequence of the FitzGerald-Lorentz contraction of length was reported in 1959 by James Terrell. Consider a cube of side A moving with uniform velocity v with respect to an observer some distance away. Figure 14-2a shows the projection of the cube on the plane containing the velocity vector **v** and the observer.

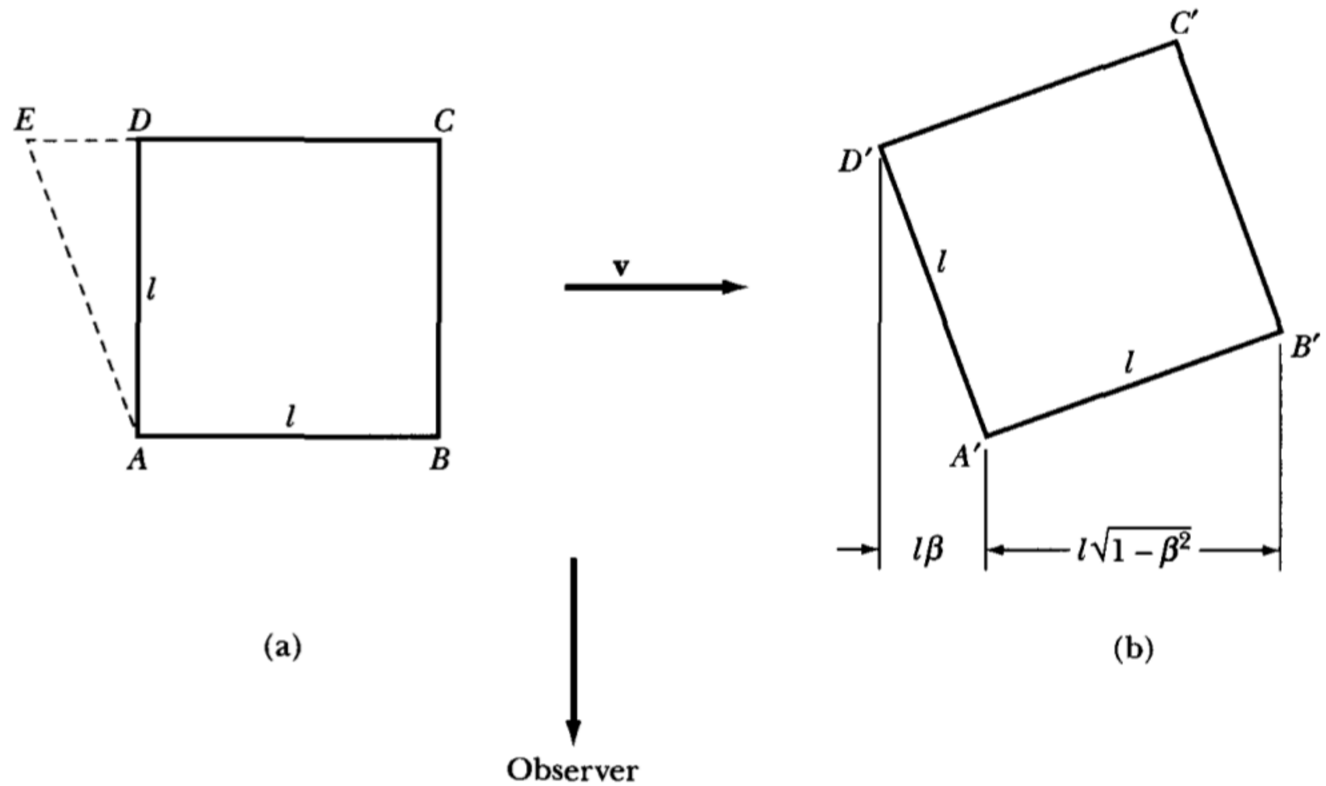


FIGURE 14-2 (a) An observer far away sees a cube of sides l at rest in system K. (b) Terrell pointed out that surprisingly, the same cube appears to be rotated if it is moving to the right with velocity **v** relative to system K.

The cube moves with its side AB perpendicular to the observer's line of sight. We wish to determine what the observer "sees"; that is, at a given instant of time in the observer's rest frame, we wish to determine the relative orientation of the corners A, B, C, and D. The traditional view (which went unquestioned for more than 50 years!) was that the only effect is a foreshortening of the sides AB and CD such that the observer sees a distorted tube of height l but of length $l\sqrt{1-\beta^2}$. Terrell pointed out that this interpretation overlooks certain facts: For light from corners A and D to reach the observer at the same instant, the light from D, which must travel a distance l farther than that from A, must have been emitted when comer D was at position E. The length DE is equal to $(l/c)v = l\beta$.

Therefore, the observer sees not only face AB, which is perpendicular to the line of sight, but also face AD, which is *parallel* to the line of sight. Also, the length of the side AB is foreshortened in the normal way to $l\sqrt{1-\beta^2}$. The net result (Figure 14-2b) corresponds exactly to the view the observer would have if the cube were rotated through an angle $\sin^{-1}\beta$. Therefore, the cube is not distorted; it undergoes an apparent rotation. Similarly, the customary statement that a moving sphere appears as an ellipsoid is incorrect; it appears still as a sphere. Computers can be used to show extremely interesting results of the type we have been discussing (Figure 14-3).

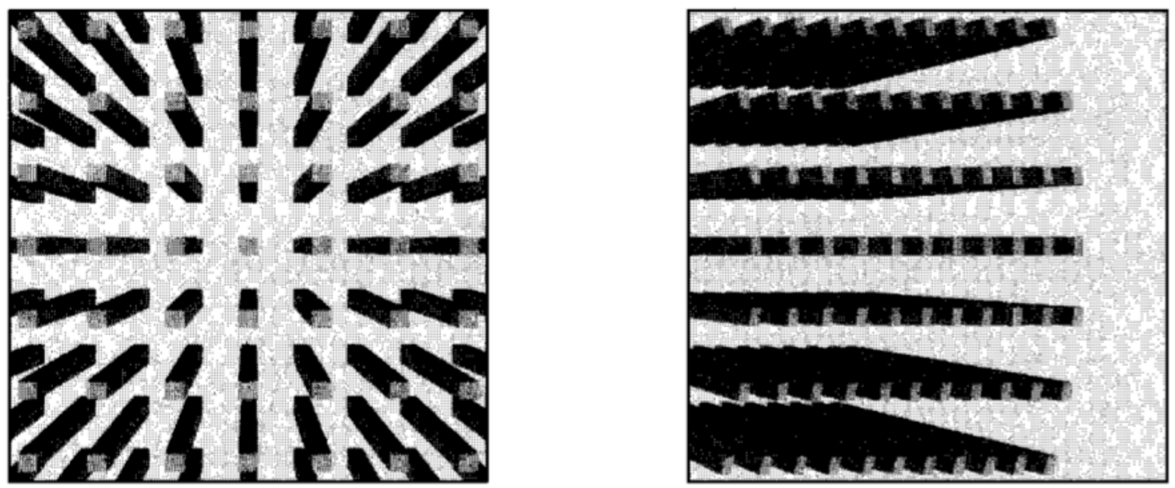


FIGURE 14-3 An array of rectangular bars is seen from above at rest in the figure on the left. In the right figure the bars are moving to the right with $v = 0.9c$. The bars appear to contract and rotate. Quoted from P.-K. Hsiung and R.H.P. Dunn, see *Science News* **137,** 232 (1990).

EXAMPLE 14.2

Use the Lorentz transformation to determine the time dilation effect.

Solution. Consider a clock fixed at a certain position (x_1) in the K system that produces signal indications with the interval

$$\Delta t = t(2) - t(1)$$

According to the Lorentz transformation (Equations 14.14), an observer in the moving system K' measures a time interval $\Delta t'$ (on the same clock) of

$$\Delta t' = t'(2) - t'(1)$$

$$= \frac{\left[t(2) - \frac{vx_1(2)}{c^2}\right] - \left[t(1) - \frac{vx_1(1)}{c^2}\right]}{\sqrt{1 - v^2/c^2}}$$

Because $x_1(2) = x_1(1)$ and because the clock is fixed in the K system, we have

$$\Delta t' = \frac{t(2) - t(1)}{\sqrt{1 - v^2/c^2}}$$

$$\boxed{\Delta t' = \frac{\Delta t}{\sqrt{1 - v^2/c^2}}} \tag{14.20}$$

Thus, to an observer in motion relative to the clock, the time intervals appear to be lengthened. This is the origin of the phrase "moving clocks run more slowly." Because the measured time interval on the moving clock is lengthened, the clock actually ticks slower. Notice that the clock is fixed in the K system, $x_1(1) = x_1(2)$, but not in the K' system, $x'_1(1) \neq x'_1(2)$.

The argument in the previous example can be reversed and the clock fixed in the K' system. The same result occurs; moving clocks run slower. The effect is called **time dilation**. It is important to note that the physical system is unimportant. The same effect occurs for a tuning fork, an hourglass, a quartz crystal, and a heartbeat. The problem is one of simultaneity. Events simultaneous in one system may not be simultaneous in another one moving with respect to the first. The same clock may be viewed from n different reference frames and found to be running at n different rates, simultaneously. Space and time are intricately interwoven. We shall return to this point later.

The time measured on a clock fixed in a system present at two events is called the **proper time** and given the symbol τ . For example, $\Delta t = \Delta\tau$ when a clock fixed in system K is present for both events, $x_1(1)$ and $x_1(2)$. Equation 14.20 becomes

$$\Delta t' = \gamma\Delta\tau \tag{14.21}$$

Notice that the proper time is always the minimum measurable time difference between two events. Moving observers always measure a longer time period

14.4 Experimental Verification of the Special Theory

The special theory of relativity explains the difficulties existing before 1900 with optics and electromagnetism. For example, the problems with stellar aberration and the Michelson-Morley experiment are solved by assuming no ether but requiring the Lorentz transformation.

But what about the new startling predictions of the special theory - length contraction and time dilation? These topics are addressed every day in the accelerator laboratories of nuclear and particle physics, where particles are accelerated to speeds close to that of light, and relativity must be considered. Other experiments can be performed with natural phenomena. We examine two of these.

Muon Decay

When cosmic rays enter the earth's outer atmosphere, they interact with particles and create cosmic showers. Many of the particles in these showers are π -mesons, which decay to other particles called muons. Muons are also unstable and decay according to the decay law, $N = N_0 e^{(-0.693t/t_{1/2})}$, where N_0 and N are the number of muons at time t = 0 and t, respectively, and $t_{1/2}$ is the half-life. However, enough muons reach the earth's surface that we can detect them easily.

Let us assume that we mount a detector on top of a 2,000-m mountain and count the number of muons traveling at a speed near v = 0.98c. Over a given period of time, we count 10^3 muons.

The half-life of muons is known to be 1.52 X 10~® s in their own rest frame (system K'). We move our detector to sea level and measure the number of muons (having v = 0.98c) detected during an equal period of time. What do we expect?

Determined classicjilly, muons traveling at a speed of 0.98c cover the 2,000 m in 6.8 X 10"® s, and 45 muons should survive the flight from 2,000 m to sea level according to the radioactive decay law. But experimental measurement indicates that 542 muons survive, a factor of 12 more.

This phenomenon must be treated relativistically. The decaying muons are moving at a high speed relative to the experimenters fixed on the earth. We therefore observe the muons' clock to be running slower. In the muons' rest frame, the time period of the muons' flight is not At = 6.8 X 10"® s but rather At/y. For v = 0.98c, y = 5, so we measure the flight time on a clock at rest in the muons' system to be 1.36 X 10"® s. The radioactive decay law predicts that 538 muons survive, much closer to our measurement and within the experimental uncertainties. An experiment similar to this has verified the time dilation prediction.

EXAMPLE 14.3

Examine the muon decay just discussed from the perspective of an observer moving with the muon.

Solution. The hjilf-life of the muon according to its own clock is 1.52 X 10^{-6} s But an observer moving with the muon would not measure the distance from the top of the mountain to sea level to be 2,000 m. According to that observer, the distance would be only 400 m. At a speed of 0.98c, it takes the muon only 1.36 X 10^{-6} s to travel the 400 m. An observer in the muon system would predict 538 muons to survive, in agreement with an observer on the earth.

Muon decay is an excellent example of a natural phenomenon that can be described in two systems moving with respect to each other. One observer sees time dilated and the other observer sees length contracted. Each, however, predicts a result in agreement with experiment.

Atomic Clock Time Measurements

An even more direct confirmation of special relativity was reported by two American physicists, J. C. Hafele and Richard E. Keating, in 1972. They used four extremely accurate cesium atomic clocks. Two clocks were flown on regularly scheduled commercial jet airplanes around the world, one eastward and one westward; the other two reference clocks stayed fixed on the earth at the U.S. Naval Observatory. A well-defined, hyperfine transition in the ground state of the ^{135}Cs atom has a frequency of 9,192,631,770 Hz and can be used as an accurate measurement of a time period.

The time measured on the two moving clocks was compared with that of the two reference clocks. The eastward trip lasted 65.4 hours with 41.2 flight hours. The westward trip, a week later, took 80.3 hours with 48.6 flight hours. The predictions are complicated by the rapid rotation of the earth and by a gravitational effect from the general theory of relativity.

We can gain some insight to the expected effect by neglecting the corrections and calculating the time difference as if the earth were not rotating. The circumference of the earth is about 4×10^7 m, and a typical jet airplane speed is almost 300 m/s. A clock fixed on the ground measures a flight time T_0 of

$$T_0 = \frac{4 \times 10^7 \text{ m}}{300 \text{ m/s}} = 1.33 \times 10^5 \text{ s} (\approx 37 \text{ hr}) \tag{14.22}$$

Because the moving clock runs more slowly, the observer on the earth would say that the moving clock measures only $T = T_0 \sqrt{1-\beta^2}$. The time difference is

$$\begin{aligned} \Delta T &= T_0 - T = T_0(1 - \sqrt{1-\beta^2}) \\ &\approx \frac{1}{2}\beta^2 T_0 \end{aligned} \tag{14.23}$$

where only the first and second terms of the power series expansion for $\sqrt{1-\beta^2}$ are kept because β^2 is so small.

$$\Delta T = \frac{1}{2}\left(\frac{300 \text{ m/s}}{3 \times 10^8 \text{ m/s}}\right)^2 (1.33 \times 10^5 \text{ s}) \tag{14.24}$$

$$= 6.65 \times 10^{-8} \text{s} = 66.5 \text{ ns}$$

This time difference is greater than the uncertainty of the measurement. Notice that in this case, the clock left on the earth actually measures more time in seconds than the moving clock. This seems at variance with our earlier comments (see Equation 14.21 and discussion). But the time period referred to in Equation 14.21 is the time between two ticks, in this case, a transition in ^{133}Cs, which we measure in seconds. It is easy to remember that moving clocks run more slowly, so that in seconds the measured time difference involves fewer ticks and, according to the definition of a second, fewer seconds.

The actual predictions and observations for the time difference are

Travel	Predicted	Observed
Eastward	-40 ± 23 ns	-59 ± 10 ns
Westward	275 ± 21 ns	273 ± 7 ns

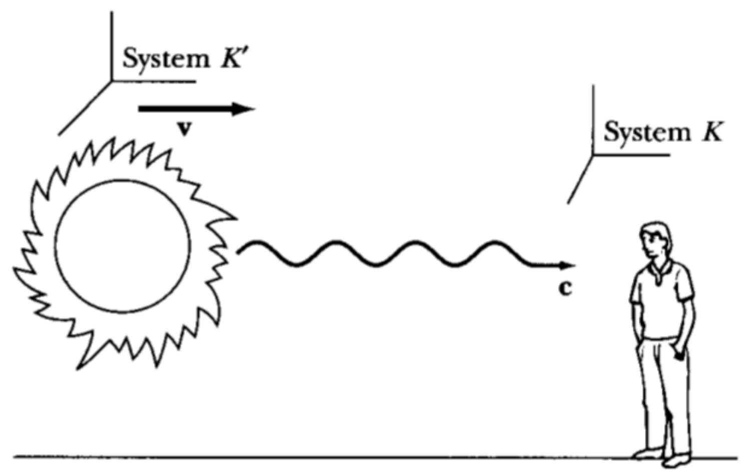
Again, the special theory of relativity is verified within the experimental uncertainties. A negative sign indicates that the time on the moving clock is less than the earth reference clock. The moving clocks lost time (ran slower) during the eastward trip and gained time (ran faster) during the westward trip. This difference is caused by the rotation of the earth, indicating that the flying clocks actually ticked faster or slower than the reference clocks on the earth. The overall positive time difference is a result of the gravitational potential effect (which we do not discuss here).

We have only briefly described two of the many experiments that have verified the special theory of relativity. There are no known experimental measurements that are inconsistent with the special theory of relativity. Einstein's work in this regard has so far withstood the test of time.

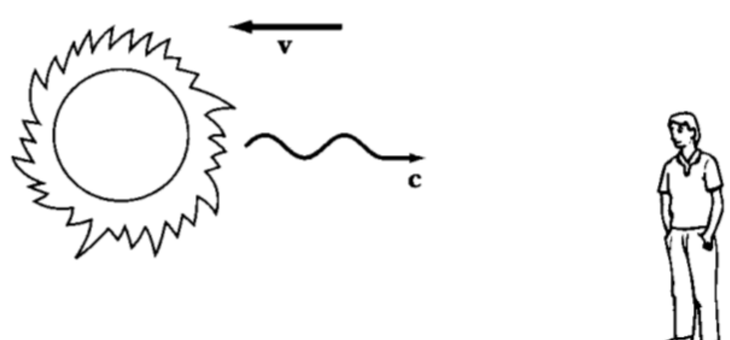
14.5 Relativistic Doppler Effect

The Doppler effect in sound is represented by an increased pitch of sound as a source approaches a receiver and a decrease of pitch as the source recedes. The change in frequency of the sound depends on whether the source or receiver is moving. This effect seems to violate Postulate I of the theory of relativity until we realize that there is a special frame for sound waves because there is a medium (e.g., air or water) in which the waves travel. In the case of light, however, there is no such medium. Only relative motion of source and receiver is meaningful in this context, and we should therefore expect some differences in the relativistic Doppler effect for light from the normal Doppler effect of sound.

Consider a source of light (e.g., a star) and a receiver approaching one another with relative speed v (Figure 14-4a).



(a) Source and receiver approaching



(b) Source and receiver receding

FIGURE 14-4 (a) An observer in system K sees light coming from a source fixed in system K'. System K' is moving toward the observer with speed v. The frequency of the light is observed in K to be increased over the value observed in K'. (b) When system K' is moving away from the observer, the frequency of the light decreases (the wavelength increases). This is the source of the term *redshifted*.

First, consider the receiver fixed in system K and the light source in system K' moving toward the receiver with speed v. During time Δt as measured by the receiver, the source emits n waves. During that time Δt, the total distance between the front and rear of the waves is

$$\text{length of wave train} = c\Delta t - v\Delta t \tag{14.25}$$

The wavelength is then

$$\lambda = \frac{c\Delta t - v\Delta t}{n} \tag{14.26}$$

and the frequency is

$$\nu = \frac{c}{\lambda} = \frac{cn}{c\Delta t - v\Delta t} \tag{14.27}$$

According to the source, it emits n waves of frequency ν_0 during the proper time $\Delta t'$:

$$n = \nu_0 \Delta t' \tag{14.28}$$

This proper time $\Delta t'$ measured on a clock in the source system is related to the time Δt measured on a clock fixed in system K of the receiver by

$$\Delta t' = \frac{\Delta t}{\gamma} \tag{14.29}$$

The clock moving with the source measures the proper time, because it is present at both the beginning and end of the waves.

Substituting Equation 14.29 into Equation 14.28, which in turn is substituted for n in Equation 14.27, gives

$$\begin{aligned}\nu &= \frac{1}{(1 - v/c)} \frac{\nu_0}{\gamma} \\ &= \frac{\sqrt{1 - v^2/c^2}}{1 - v/c} \nu_0\end{aligned} \tag{14.30}$$

which can be written as

$$\nu = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}}\nu_0 \quad \text{source and receiver approaching} \tag{14.31}$$

Equation 14.31 is also valid when the source is fixed and the receiver approaches it with speed v. Next, we consider the case in which the source and receiver recede from each other with velocity v (Figure 14-4b). The derivation is similar to the one just presented—with one small exception. In Equation 14.25, the distance between the beginning and end of the waves becomes

$$\text{length of wave train} = c\Delta t + v\Delta t \tag{14.32}$$

This change in sign is propagated through Equations 14.30 and 14.31, giving

$$\nu = \frac{\sqrt{1 - v^2/c^2}}{1 + v/c}\nu_0$$

$$\nu = \frac{\sqrt{1-\beta}}{\sqrt{1+\beta}}\nu_0 \quad \text{source and receiver receding} \tag{14.33}$$

Equations 14.31 and 14.33 can be combined into one equation.

$$\nu = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}}\nu_0 \quad \text{relativistic Doppler effect} \tag{14.34}$$

if we agree to use a + sign for β (+v/ c) when the source and receiver are approaching each other and a - sign for β when they are receding.

The relativistic Doppler effect is important in astronomy. Equation 14.34 indicates that, if the source is receding at high speed from an observer, then a lower frequency (or longer wavelength) is observed for certain spectral lines or characteristic frequencies. This is the origin of the term red shift, the wavelengths of visible light are shifted toward longer wavelengths (red) if the source is receding from us.

Astronomical observations indicate that the universe is expanding. The farther away a star is, the faster it appears to be moving away (or the greater its red shift). These data are consistent with the "big bang" origin of the universe, which is estimated to have occurred some 13 billion years ago.

EXAMPLE 14.4

During a spaceflight to a distant star, an astronaut and her twin brother on the earth send radio signals to each other at annual intervals. What is the frequency of the radio signals each twin receives from the other during the flight to the star if the astronaut is moving at v = 0.8 c? What is the frequency during the return flight at the same speed?

Solution. We use Equation 14.34 to determine the frequency of radio signals that each receives from the other. The frequency ν_0 =1 signal/year. On the leg of the trip away from the earth, β = -0.8 and Equation 14.34 gives

$$\nu = \frac{\sqrt{1 - 0.8}}{\sqrt{1 + 0.8}} \nu_0 = \frac{\nu_0}{3}$$

The radio signals are received once every 3 years.

On the return trip, however, β = +0.8 and Equation 14.34 gives $\nu = 3\nu_0$, so the radio signals are received every 4 months. In this way, the twin on the earth can monitor the progress of his astronaut twin.

14.6 Twin Paradox

Consider twins who choose different career paths. Mary becomes an astronaut, and Frank decides to be a stockbroker At age 30, Mary leaves on a mission to a planet in a nearby star's system. Mary will have to travel at a high speed to reach the planet and return.

According to Frank, Mary's biological clock will tick more slowly during her trip, so she will age more slowly. He expects Mary to look and appear younger than he does when she returns. According to Mary, however, Frank will appear to be moving rapidly with respect to her system, and she thinks Frank will be younger when she returns. This is the paradox. Which twin, if either, is younger when Mary (the moving twin) returns to the earth where Frank (the fixed twin) has remained? Because the two expectations are so contradictory, doesn't Nature have a way to prove they will be the same age?

This paradox has existed almost since Einstein first published his special theory of relativity. Variations of the argument have been presented many times. The correct answer is that Mary, the astronaut, will return younger than her twin brother, Frank, who remains busy on Wall Street. The correct analysis is as follows. According to Frank, Mary's spaceship blasts off and quickly reaches a coasting speed of $v = 0.8c$, travels a distance of 8 ly (ly = a light year, the distance light travels in 1 year) to the planet, and quickly decelerates for a short visit to the planet. The acceleration and deceleration times are negligible compared with the total travel time of 10 years to the planet. The return trip also takes 10 years, so on Mary's return to Earth, Frank will be 30 + 10 + 10 = 50 years old. Frank calculates that Mary's clock is ticking slower and that each leg of the trip takes only $10\sqrt{1 - 0.8^2} = 6$ years. Mary therefore is only 30 + 6 + 6 = 42 years old when she returns. Frank's clock is (almost) in an inertial system.

When Mary performs the time measurements on her clock, they may be invalid according to the special theory because her system is not in an inertial frame of reference moving at a constant speed with respect to the earth. She accelerates and decelerates at both the earth and the planet, and to make valid time measurements to compare with Frank's clock, she must account for this acceleration and deceleration. The instantaneous rate of Mary's clock is still given by Equation 14.20, because the instantaneous rate is determined by the instantaneous speed v. Thus, there is no paradox if we obey the two postulates of the special theory. It is also clear which twin is in the inertial frame of reference. Mary will actually feel the forces of acceleration and deceleration. Frank feels no such forces. When Mary returns home, her twin brother has invested her 20 years of salary, making her a rich woman at the young age of 42. She was paid a 20-year salary for a job that took her only 12 years!

EXAMPLE 14.5

Mary and Frank send radio signals to each other at 1-year intervals after she leaves Earth. Analyze the times of receipt of the radio messages. Solution. In Example 14.4, we calculated that such radio signals are received every 3 years on the trip out and every $\frac{1}{3}$ year on the trip back. First, we examine the signals Mary receives from Frank. During the 6-year trip to the planet, Mary receives only two radio messages, but on the 6-year return trip, she receives eighteen signals, so she correctly concludes that her twin brother Frank has aged 20 years and is now 50 years old.

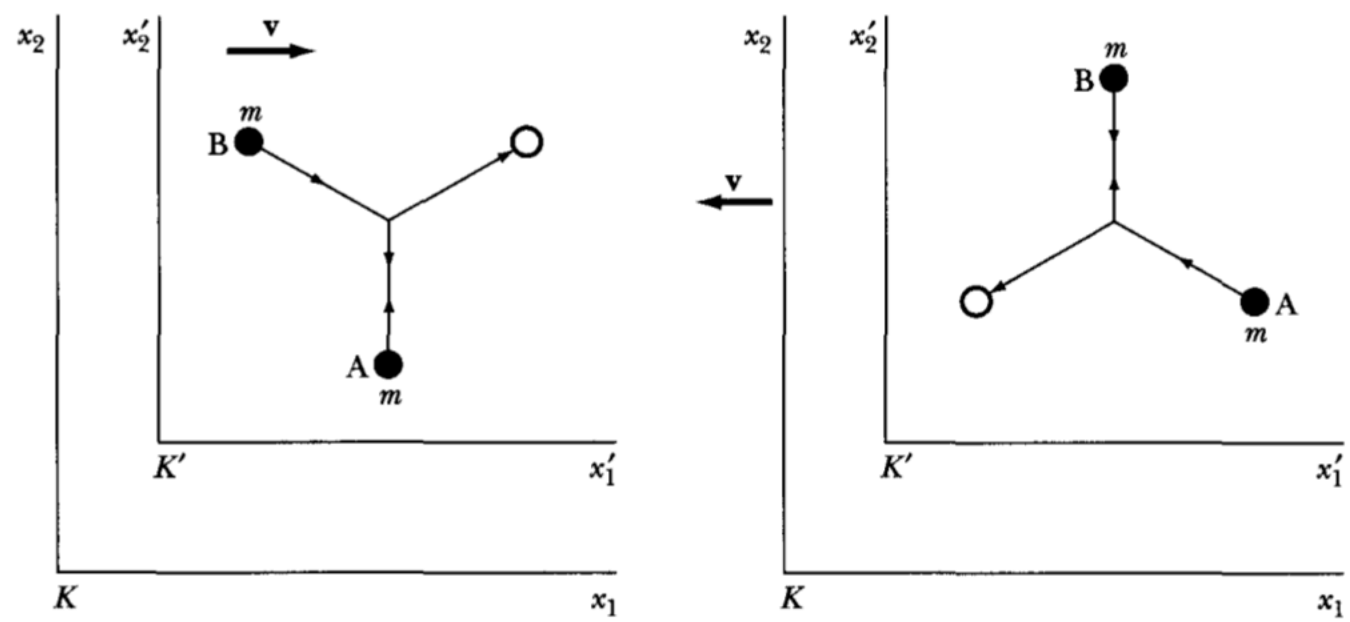
In Frank's system, Mary's trip to the planet takes 10 years. By the time Mary reaches the planet, Frank receives 10/3 signals (i.e., three signals plus one-third of the time to the next one). However, Frank continues to receive a signal every 3 years for the 8 years it takes the last signal Mary sends when she reaches the planet to travel to Frank. Thus, Frank receives signals every 3 years for 8 more years (total of 18 years) for a total of six radio signals from the period of travel to the planet. Frank has no way of knowing that Mary has stopped and turned around until the radio message, which takes 8 years, is received. Of the remaining 2 years of Mary's journey according to Frank (20 - 18 = 2), Frank receives signals every $\frac{1}{3}$ year, or six more signals. Frank correctly determines that Mary has aged 6 + 6 = 12 years during her journey because he receives a total of 12 signals.

Thus, both twins agree about their own ages and about each other's. Mary is 42 and Frank is 50 years old.

14.7 Relativistic Momentum

Newton's Second Law, $\mathbf{F} = d\mathbf{p}/dt$, is covariant under a Galilean transformation. Therefore, we do not expect it to keep its form under a Lorentz transformation. We can foresee difficulties with Newton's laws and the conservation laws unless we make some necessary changes. According to Newton's Second Law, for example, an acceleration at high speeds might cause a particle's velocity to exceed c, an impossible condition according to the special theory of relativity.

We begin by examining the conservation of linear momentum in a force-free (no external forces) collision. There are no accelerations. Observer A at rest in system K holds a ball of mass m, as does observer B in system K' moving to the right with relative speed v with respect to system K, as in Figure 14-1. The two observers throw their (identical) balls along their respective x_2-axes, which results in a perfectly elastic collision. The collision, according to observers in the two systems, is shown in Figure 14-5. Each observer measures the speed of his or her ball to be u_0.



(a) Collision according to system K (b) Collision according to system K'

FIGURE 14-5 Observer A, at rest in fixed system K, throws a ball straight up in system K. Observer B, at rest in system K', which is moving to the right with velocity $\mathbf{v}$, throws a ball straight down so that the two balls collide. (a) The collision according to observer A in system K. (b) The collision according to observer B in system K'. Each observer measures the speed of his or her ball to be u_0. We examine the linear momentum of the ball.

We first examine the conservation of momentum according to system K. The velocity of the ball thrown by observer A has components

$$\left.\begin{aligned} u_{A1} &= 0 \\ u_{A2} &= u_0 \end{aligned}\right\} \tag{14.35}$$

The momentum of ball A is in the x_2-direction:

$$p_{A2} = mu_0 \tag{14.36}$$

The collision is perfectly elastic, so the ball returns down with speed u_0. The change in momentum observed in system K is

$$\Delta p_{A2} = -2mu_0 \tag{14.37}$$

Does Equation 14.37 also represent the change in momentum of the ball thrown by observer B in the moving system K'? We use the inverse velocity transformation of Equations 14.17 (i.e., we interchange primes and unprimes and let $v \to -v$) to determine

$$\left.\begin{aligned} u_{B1} &= v \\ u_{B2} &= -u_0\sqrt{1 - v^2/c^2} \end{aligned}\right\} \tag{14.38}$$

where $u'_{B1} = 0$ and $u'_{B2} = u_0$. The momentum of ball B and its change in momentum during the collision become

$$p_{B2} = -mu_0\sqrt{1 - v^2/c^2} \tag{14.39}$$

$$\Delta p_{B2} = +2mu_0\sqrt{1 - v^2/c^2} \tag{14.40}$$

Equations 14.37 and 14.40 do not add to zero: *Linear momentum is not conserved according to the special theory if we use the conventions for momentum of classical physics*. Rather than abandoning the law of conservation of momentum, we look for a solution that allows us to retain both it and Newton's Second Law.

As we did for the Lorentz transformation, we assume the simplest possible change. We assume that the classical form of momentum mu is multiplied by a constant that may depend on speed k(u):

$$\mathbf{p} = k(u)m\mathbf{u} \tag{14.41}$$

In Example 14.6, we show that the value

$$k(u) = \frac{1}{\sqrt{1 - u^2/c^2}} \tag{14.42}$$

allows us to retain the conservation of linear momentum. Notice that the form of Equation 14.42 is the same as that found for the Lorentz transformation. In fact, the constant k(u) is given the same label: λ . However, this λ contains the speed of the particle u, whereas the Lorentz transformation contains the relative speed v between the two inertial reference frames. This distinction must be kept in mind; it often causes confusion.

We can make a plausible calculation for the relativistic momentum if we use the proper time τ (see Equation 14.21) rather than the normal time t. In this case

$$\mathbf{p} = m\frac{d\mathbf{x}}{d\tau} = m\frac{d\mathbf{x}}{dt}\frac{dt}{d\tau} \tag{14.43}$$

$$= m\frac{d\mathbf{x}}{dt}\frac{1}{\sqrt{1 - u^2/c^2}} \tag{14.44}$$

$$\boxed{\mathbf{p} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} = \gamma m\mathbf{u}} \quad \textbf{relativistic momentum} \tag{14.45}$$

where we retain **u** = d**x**/dt as used classically.

Although all observers do not agree as to d**x**/dt, they do agree as to d**x**/dτ , where the proper time dτ is measured by the moving object itself. The relation dt/dτ is obtained from Equation 14.21, where the speed u has been used in γ to represent the speed of a reference frame fixed in the object that is moving with respect to a fixed frame.

Equation 14.45 is our new definition of momentum, called **relativistic momentum**. Notice that it reduces to the classical result for small values of u/c. It was fashionable in past years to call the mass in Equation 14.45 the **rest mass** m_0 and to call the term

$$m = \frac{m_0}{\sqrt{1 - u^2/c^2}} \quad \textbf{(old-fashioned notation)} \tag{14.46}$$

the **relativistic mass**. The term rest mass resulted from Equation 14.46 when u = 0, and the classical form of momentum was thus retained: **p** = m**u**. Scientists spoke of the mass increasing at high speeds. We prefer to keep the concept of mass as an invariant, intrinsic property of an object. The use of the two terms *relativistic* and *rest mass* is now considered old-fashioned, although the terms are still sometimes used. *We always refer to the mass m, which is the same as the rest mass*. The use of relativistic mass often leads to mistakes when using classical expressions.

EXAMPLE 14.6

Show that linear momentum is conserved in the x_2-direction for the collision shown in Figure 14-5 if relativistic momentum is used.

Solution. We can modify the classical expressions for momentum already obtained for the two balls. The momentum for ball A becomes (from Equation 14.36)

$$p_{A2} = \frac{mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.47}$$

and

$$\Delta p_{A2} = \frac{-2mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.48}$$

Before modifying Equation 14.39 for the momentum of ball B, we must first find the speed of ball B as measured in system K We use Equation 14.38 to determine

$$\begin{aligned} u_B &= \sqrt{u_{B1}^2 + u_{B2}^2} \\ &= \sqrt{v^2 + u_0^2(1 - v^2/c^2)} \end{aligned} \tag{14.49}$$

The momentum p_{B2} is found by modifying Equation 14.39:

$$p_{B2} = -mu_0\gamma\sqrt{1 - v^2/c^2}$$

where

$$\gamma = \frac{1}{\sqrt{1 - u_B^2/c^2}}$$

$$p_{B2} = \frac{-mu_0\sqrt{1 - v^2/c^2}}{\sqrt{1 - u_B^2/c^2}} \tag{14.50}$$

Using u_B from Equation 14.49 gives

$$p_{B2} = \frac{-mu_0\sqrt{1 - v^2/c^2}}{\sqrt{(1 - u_0^2/c^2)(1 - v^2/c^2)}}$$

$$= \frac{-mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.51}$$

$$\Delta p_{B2} = \frac{+2mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.52}$$

Equations 14.48 and 14.52 add to zero, as required for the conservation of linear momentum.

14.8 Energy

With a new definition of linear momentum (Equation 14.45) in hand, we turn our attention to energy and force. We keep our former definition (Equation 2.86) of kinetic energy as being the work done on a particle. The work done is defined in Equation 2.84 to be

$$W_{12} = \int_1^2 \mathbf{F} \cdot d\mathbf{r} = T_2 - T_1 \tag{14.53}$$

Equation 2.2 for Newton's Second Law is modified to account for the new definition of linear momentum:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(\gamma m\mathbf{u}) \tag{14.54}$$

If we start from rest, $T_1 = 0$, and the velocity **u** is initially along the direction of the force.

$$W = T = \int \frac{d}{dt}(\gamma m\mathbf{u}) \cdot \mathbf{u}\, dt \tag{14.55}$$

$$= m\int_0^u u\, d(\gamma u) \tag{14.56}$$

Equation 14.56 is integrated by parts to obtain

$$T = \gamma mu^2 - m\int_0^u \frac{udu}{\sqrt{1 - u^2/c^2}}$$

$$= \gamma mu^2 + mc^2\sqrt{1 - u^2/c^2}\,\Big|_0^u$$

$$= \gamma mu^2 + mc^2\sqrt{1 - u^2/c^2} - mc^2 \tag{14.57}$$

With algebraic manipulation. Equation 14.57 becomes

$$\boxed{T = \gamma mc^2 - mc^2} \quad \textbf{relativistic kinetic energy} \tag{14.58}$$

Equation 14.58 seems to resemble in no way our former result for kinetic energy, $T = \frac{1}{2}mu^2$. However, Equation 14.58 must reduce to $\frac{1}{2}mu^2$ for small values of velocity.

EXAMPLE 14.7

Show that Equation 14.58 reduces to the classical result for small speeds, $u \ll c$.

Solution. The first term of Equation 14.58 can be expanded in a power series:

$$T = mc^2 + \frac{1}{2}mu^2 - mc^2$$
$$= \frac{1}{2}mu^2 \qquad (14.59)$$

where all terms of power $(u/c)^4$ or greater are neglected because $u << c$.

$$T = mc^2 + \frac{1}{2}mu^2 - mc^2$$
$$= \frac{1}{2}mu^2 \qquad (14.60)$$

which is the classical result.

It is important to note that neither $\frac{1}{2}mu^2$ nor $\frac{1}{2}mu^2$ gives the correct relativistic value for the kinetic energy.

The term mc^2 in Equation 14.58 is called the **rest energy** and is denoted by Eq.

$$\boxed{E_0 \equiv mc^2} \quad \text{rest energy} \qquad (14.61)$$

Equation 14.58 is rewritten

$$\gamma mc^2 = T + mc^2$$

Thus,

$$E = T + E_0 \qquad (14.62)$$

where

$$\boxed{E \equiv \gamma mc^2 = T + E_0} \quad \text{total energy} \qquad (14.63)$$

The total energy, $E = \gamma mc^2$, is defined as the sum of kinetic energy and the rest energy. Equations 14.58-14.63 are the origin of Einstein's famous relativistic result of the equivalence of mass and energy (energy = mc^2). These equations are consistent with this interpretation. Note that when a body is not in motion ($u = 0 = T$), Equation 14.63 indicates that the total energy is equal to the rest energy

If mass is simply another form of energy, then we must combine the classical conservation laws of mass and energy into one conservation law of mass-energy represented by Equation 14.63. This law is easily demonstrated in the atomic nucleus, where the mass of constituent particles is converted to the energy that binds the individual particles together.

EXAMPLE 14.8

Use the atomic masses of the particles involved to calculate the binding energy of a deuteron.

Solution. A deuteron is composed of a neutron and a proton. We use atomic masses, because the electron masses cancel.

$$
\begin{aligned}
\text{mass of neutron} &= 1.008665\ \text{u} \\
\text{mass of proton } (^1\text{H}) &= \underline{1.007825\ \text{u}} \\
\text{sum} &= 2.016490\ \text{u} \\
\text{mass of deuteron } (^2\text{H}) &= 2.014102\ \text{u} \\
\text{difference} &= 0.002388\ \text{u}
\end{aligned}
$$

This difference in mass-energy is equal to the binding energy holding the neutron and proton together as a deuteron. The mass units are atomic mass units (u), which can be converted to kilograms if necessary. However, the conversion of mass to energy is facilitated by the well-known relation between mass and energy:

$$1\ \text{u}c^2 = 931.5\ \text{MeV} \tag{14.64}$$

The binding energy of the deuteron is therefore

$$0.002388\ \text{u}c^2 \times 931.5 \frac{\text{MeV}}{\text{u}c^2} = 2.22\ \text{MeV}$$

Nuclear experiments of the form $\gamma + 2\text{H} \rightarrow \text{n} + \text{p}$ indicate that gamma rays of energy just greater than 2.22 MeV are required to break the deuteron apart into a neutron and a proton. Conversely, when a neutron and proton join at rest to form a deuteron, 2.22 MeV of energy is released in the form of kinetic energy

Because physicists believe that momentum is a more fundamental concept than kinetic energy (for example, there is no general law of conservation of kinetic energy), we would like a relation for mass-energy that includes momentum rather than kinetic energy. We begin with Equation 14.45 for momentum:

$$p = \gamma m u$$

$$p^2c^2 = \gamma^2 m^2 u^2 c^2$$

$$= \gamma^2 m^2 c^4 \left(\frac{u^2}{c^2}\right) \tag{14.65}$$

It is easy to show that

$$\frac{u^2}{c^2} = 1 - \frac{1}{\gamma^2} \tag{14.66}$$

so Equation 14.65 becomes

$$p^2c^2 = \gamma^2 m^2 c^4 \left(1 - \frac{1}{\gamma^2}\right)$$

$$= \gamma^2 m^2 c^4 - m^2 c^4$$

$$= E^2 - E_0^2$$

$$\boxed{E^2 = p^2c^2 + E_0^2} \tag{14.67}$$

Equation 14.67 is a very useful kinematic relationship. It relates the total energy of a particle to its momentum and rest energy.

Notice that a photon has no mass, so that Equation 14.67 gives

$$E = pc \quad \text{photon} \tag{14.68}$$

There is no such thing as a photon at rest

14.9 Spacetime and Four-Vectors

In Section 14.3 (Equation 14.5), we noticed that the quantities

$$\left.\begin{aligned}\sum_{j=1}^{3} x_j^2 - c^2t^2 = 0\\ \sum_{j=1}^{3} x_j'^2 - c^2t'^2 = 0\end{aligned}\right\}$$

are invariant because the speed of light is the same in all inertial systems in relative motion. Consider two events separated by space and time. In system K,

$$\Delta x_i = x_i(\text{event 2}) - x_i(\text{event 1})$$

$$\Delta t = t(\text{event 2}) - t(\text{event 1})$$

The interval Δs^2 is invariant in all inertial systems in relative motion

$$\Delta s^2 = \sum_{j=1}^{3} (\Delta x_j)^2 - c^2\Delta t^2 \tag{14.69}$$

$$\Delta s^2 = \Delta s'^2 = \sum_{j=1}^{3} (\Delta x_j')^2 - c^2\Delta t'^2 \tag{14.70}$$

Equation 14.69 can be written as a differential equation:

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 - c^2dt^2 \tag{14.71}$$

Consider the system K', where the particle is instantaneously at rest. Because $dx'_1 = dx'_2 = dx'_3 = 0$ in this case, $dt' = d\tau$, the proper time interval discussed above (Equation 14.21). Equation 14.70 becomes

$$-c^2d\tau^2 = dx_1^2 + dx_2^2 + dx_3^2 - c^2dt^2 \tag{14.72}$$

Using the Lorentz transformation, Equation 14.72 gives a similar result to Equation 14.21:

$$d\tau = \frac{dt}{\gamma} \tag{14.73}$$

The proper time τ is, along with the length quantity Δs^2, another Lorentz invariant quantity.

A useful concept in special relativity is that of the **light cone**. The invariant length Δs^2 suggests adding ct as a fourth dimension to the three space dimensions x_1, x_2, and x_3. In Figure 14-6, we plot ct versus one of the Euclidean space coordinates. The origin of (x, ct) is the present (0, 0). The solid lines represent the paths taken in the past and in the future by light. A particle traveling the path from A to B is said to be moving along its **worldline**. For time $t < 0$, the particle has been in the lower cone, the past. Similarly, for $t > 0$ the particle will move in the upper cone, the future. It is not possible for us to know about events outside the light cone; this region, called "elsewhere," requires $v > c$.

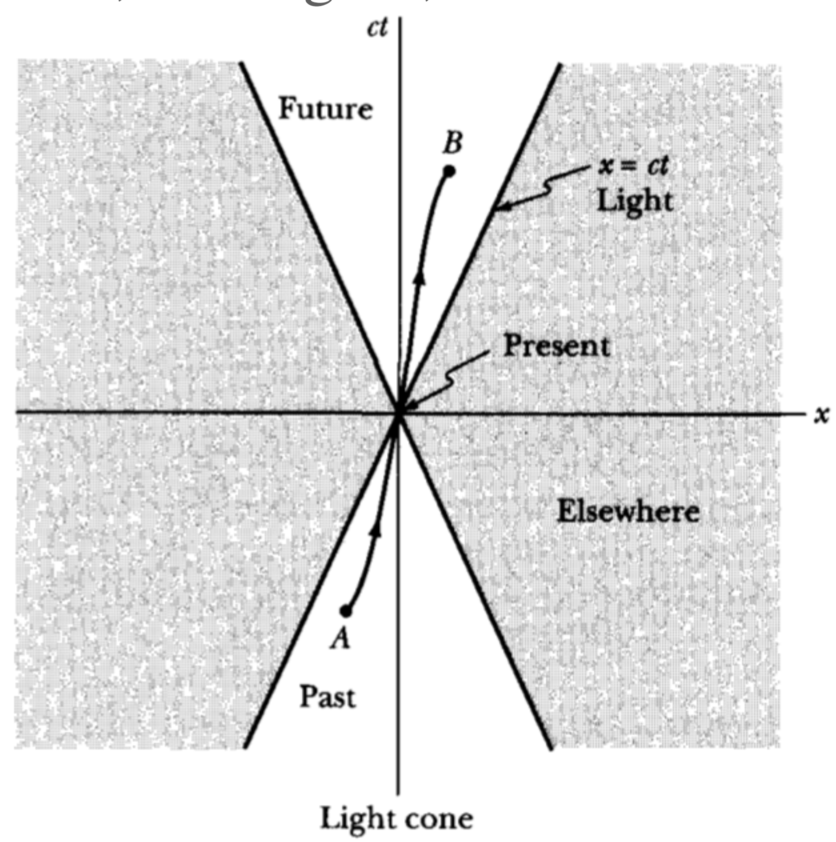


FIGURE 14-6 The variable *ct* is plotted versus *x* with the origin being the present. The heavy solid lines indicate the past and future paths of light and form a *light cone.* To the right and left of these lines is considered "elsewhere," because we cannot reach this region from the present. The path from *A* to *B* represents a *worldline,* a path that we can take traveling at speeds less than or equal to light.

There are two possibilities concerning the value of Δs^2. If $\Delta s^2 > 0$, the two events have a spacelike interval. One can always find an inertial frame traveling with v<csuch that the two events occur at different space coordinates but at the same time. When $\Delta s^2 < 0$, the two events are said to have a timelike interval. One can always find a suitable inertial frame in which the events occur at the same point in space but at different times. In the case $\Delta s^2 = 0$, the two events are separated by a light ray.

Only events separated by a timelike interval can be causally connected. The present event in the light cone can be causally related only to events in the past region of the light cone. Events with a spacelike interval cannot be causally connected. Space and time, although distinct, are nonetheless intricately related.

The previous discussion of space and time suggests using ct as a fourth dimensional parameter We continue this line of thought by defining $x_4 = ict$ and $x'_4 = ict'$. The use of the imaginary number i does not indicate that this component is imaginary. The imaginary number simply allows us to represent the relations in concise, mathematical form.

The rest of this section could just as well be carried out without the use of i(e.g., $x_4 = ct$), but the mathematics would be more cumbersome. The useful results are in terms of real, physical quantities.

Using $x_4 = ict$ and $x'_4 = ict'$, we can write Equations 14.5 as

$$\left.\begin{aligned} \sum_{\mu=1}^{4} x_\mu^2 &= 0 \\ \sum_{\mu=1}^{4} x_\mu'^2 &= 0 \end{aligned}\right\} \qquad (14.74)$$

In accordance with standard convention, we use Greek indices (usually μ or ν) to indicate summations that run from 1 to 4; in relativity theory, Latin indices are usually reserved for summations that run from 1 to 3.

From these equations, it is clear that the two sums must be proportional(proof is given in Appendix G), and because the motion is symmetrical between the systems, the proportionality constant is unity. Thus,

$$\sum_{\mu} x_{\mu}^2 = \sum_{\mu} x_{\mu}'^2 \tag{14.75}$$

This relation is analogous to the three-dimensional, distance-preserving, orthogonal rotations we have studied previously (see Section 1.4) and indicates that the Lorentz transformation corresponds to a rotation in a four-dimensional space (called **world space** or **Minkowski space**). The Lorentz transformations are then orthogonal transformations in Minkowski space:

$$x_{\mu}' = \sum_{\nu} \lambda_{\mu\nu} x_{\nu} \tag{14.76}$$

where the $\lambda_{\mu\nu}$ are the elements of the Lorentz transformation matrix. From Equations 14.14, the transformation λ is

$$\boldsymbol{\lambda} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \tag{14.77}$$

A quantity is called a **four-vector** if it consists of four components, each of which transforms according to the relation

$$A_{\mu}' = \sum_{\nu} \lambda_{\mu\nu} A_{\nu} \tag{14.78}$$

where the $\lambda_{\mu\nu}$ define a Lorentz transformation. Such a four-vector(openface capital letters) is

$$\mathbb{X} = (x_1, x_2, x_3, ict) \tag{14.79a}$$

or

$$\boxed{\mathbb{X} = (\mathbf{x}, ict)} \tag{14.79b}$$

where the notation of the last line means that the first three (space) components of $\mathbb{X}$ define the ordinary three-dimensional position vector **x** and that the fourth component is ict. Similarly, the differential of $\mathbb{X}$ is a four-vector:

$$d\mathbb{X} = (d\mathbf{x},\ ic\ dt) \tag{14.80}$$

In Minkowski space, the four-dimensional element of length is invariant. Its magnitude is unaffected by a Lorentz transformation, and such a quantity is called a four-scalar or world scalar. Equation 14.71 can be written as

$$ds = \sqrt{\sum_{\mu} dx_\mu^2} \tag{14.81}$$

and Equation 14.72 as

$$d\tau = \frac{i}{c}\sqrt{\sum_{\mu} dx_\mu^2} = \frac{i}{c}\,ds \tag{14.82}$$

The proper time dr is invariant because it is simply i/ c times the element of The proper time dτ is invariant because it is simply i/ c times the element of length ds. The ratio of the four-vector $d\mathbb{X}$ to the invariant dτ is therefore also a four-vector, called the four-vector velocity $\mathbb{V}$:

$$\boxed{\mathbb{V} = \frac{d\mathbb{X}}{d\tau} = \left(\frac{d\mathbf{x}}{d\tau},\ ic\frac{dt}{d\tau}\right)} \tag{14.83}$$

The components of the ordinary velocity u are

$$u_j = \frac{dx_j}{dt}$$

so, using Equations 14.71 and 14.82, dτ can be expressed as

$$d\tau = dt\sqrt{1 - \frac{1}{c^2}\sum_j \frac{dx_j^2}{dt^2}}$$

or

$$d\tau = dt\sqrt{1-\beta^2} \tag{14.84}$$

as we found in Equation 14.73. The four-vector velocity can therefore be written as

$$\boxed{\mathbb{V} = \frac{1}{\sqrt{1-\beta^2}}(\mathbf{u}, ic)} \tag{14.85}$$

where **u** represents the three space components of ordinary velocity, u_1, u_2, u_3. (Remember that the particle's velocity is now denoted by **u** to distinguish it from the moving frame velocity **v**.) The four-vector momentum is now simply the mass times four-vector velocity, because mass is invariant;

$$\mathbb{P} = m\mathbb{V} \tag{14.86}$$

$$\boxed{\mathbb{P} = \left(\frac{m\mathbf{u}}{\sqrt{1-\beta^2}}, ip_4\right)} \tag{14.87}$$

where

$$p_4 \equiv \frac{mc}{\sqrt{1-\beta^2}} \tag{14.88}$$

The first three components of the four-vector momentum $\mathbb{P}$ are the components of the relativistic momentum (Equation 14.45):

$$P_j = p_j = \gamma m u_j, \quad j = 1, 2, 3 \tag{14.89}$$

Using Equation 14.63, the fourth component of the momentum is related to the total energy E:

$$p_4 = \gamma mc = \frac{E}{c} \tag{14.90}$$

The four-vector momentum can therefore be written as

$$\mathbb{P} = \left(\mathbf{p}, i\frac{E}{c}\right) \tag{14.91}$$

where **p** stands for the three space components of momentum. Thus, in relativity theory, momentum and energy are linked in a manner similar to that which joins the concepts of space and time. If we apply the Lorentz transformation matrix (Equation 14.77) to the momentum $\mathbb{P}$, we find

$$\boxed{\begin{aligned} p_1' &= \frac{p_1 - (v/c^2)E}{\sqrt{1-\beta^2}} \\ p_2' &= p_2 \\ p_3' &= p_3 \\ E' &= \frac{E - vp_1}{\sqrt{1-\beta^2}} \end{aligned}} \tag{14.92}$$

EXAMPLE 14.9

Using the methods of this section, derive Equation 14.67.

Solution. If we place the origin of the moving system K' fixed on the particle, we have u = v. The square of the four-vector velocity (Equation 14.85) is invariant:

$$\mathbb{V}^2 = \sum_\mu V_\mu^2 = \frac{v^2 - c^2}{1-\beta^2} = -c^2 \tag{14.93}$$

Hence, the square of the four-vector momentum is also invariant:

$$\mathbb{P}^2 = \sum_\mu P_\mu^2 = m^2\mathbb{V}^2 = -m^2c^2 \tag{14.94}$$

From Equation 14.91, we also have, using $\mathbf{p}\cdot\mathbf{p} = p^2 = p_1^2 + p_2^2 + p_3^2$

$$\mathbb{P}^2 = p^2 - \frac{E^2}{c^2} \tag{14.95}$$

Combining the last two equations gives Equation 14.67.

$$E^2 = p^2c^2 + m^2c^4 = p^2c^2 + E_0^2$$

If we define an angle ϕ such that $\beta = \sin\phi$, the relativistic relations between velocity, momentum, and energy can be obtained by trigonometric relations involving the so-called "relativistic triangle" (Figure 14-7).

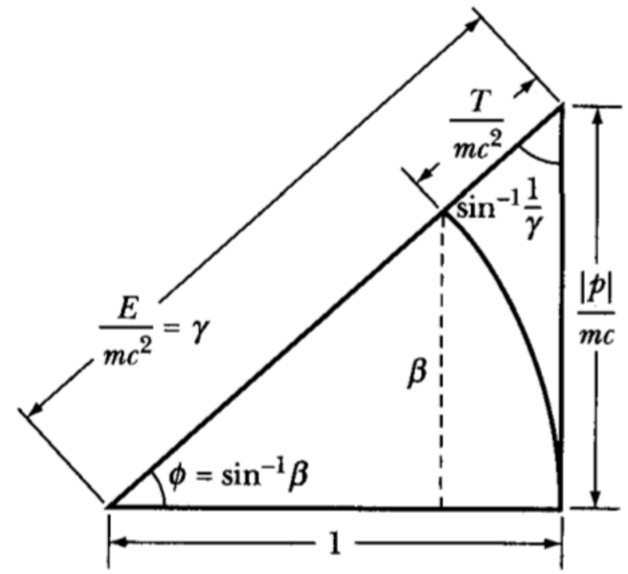


FIGURE 14-7 The relativistic triangle allows us to find relations between velocity, momentum, and energy by using trigonometric relations.

EXAMPLE 14.10

Derive the velocity addition rule.

Solution. Suppose that there are three inertial reference frames, K, K', and K", which are in collinear motion along their respective x_1-axes. Let the velocity of K' relative to K be v_1 and let the velocity of K" relative to K' be v_2. The speed of K" relative to K cannot be $v_1 + v_2$, because it must be possible to propagate a signal between any two inertial frames, and if both v_1 and v_2 are greater than c/2 (but less than c), then $v_1 + v_2 > c$. Therefore, the rule for the addition of velocities in relativity must be different from that in Galilean theory. The relativistic velocity addition rule can be obtained by considering the Lorentz transformation matrix connecting K and K". The individual transformation matrices are

$$\boldsymbol{\lambda}_{K'\to K} = \begin{pmatrix} \gamma_1 & 0 & 0 & i\beta_1\gamma_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta_1\gamma_1 & 0 & 0 & \gamma_1 \end{pmatrix}$$

$$\boldsymbol{\lambda}_{K''\to K'} = \begin{pmatrix} \gamma_2 & 0 & 0 & i\beta_2\gamma_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta_2\gamma_2 & 0 & 0 & \gamma_2 \end{pmatrix}$$

The transformation from K" to K is just the product of these two transformations:

$$\boldsymbol{\lambda}_{K''\to K} = \boldsymbol{\lambda}_{K''\to K'}\boldsymbol{\lambda}_{K'\to K} = \begin{pmatrix} \gamma_1\gamma_2(1+\beta_1\beta_2) & 0 & 0 & i\gamma_1\gamma_2(\beta_1+\beta_2) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\gamma_1\gamma_2(\beta_1+\beta_2) & 0 & 0 & \gamma_1\gamma_2(1+\beta_1\beta_2) \end{pmatrix}$$

So that the elements of this matrix correspond to those of the normal Lorentz matrix (Equation 14.77), we must identify β and γ for the K" —> K transformation as

$$\left.\begin{aligned} \gamma &= \gamma_1\gamma_2(1+\beta_1\beta_2) \\ \beta\gamma &= \gamma_1\gamma_2(\beta_1+\beta_2) \end{aligned}\right\} \tag{14.96}$$

from which we obtain

$$\beta = \frac{\beta_1+\beta_2}{1+\beta_1\beta_2} \tag{14.97}$$

If we multiply this last expression by c, we have the usual form of the velocity (speed) addition rule:

$$\boxed{v = \frac{v_1+v_2}{1+(v_1v_2/c^2)}} \tag{14.98}$$

It follows that if $v_1 < c$ and $v_2 < c$, then $v < c$ also.

Even though *signal* velocities can never exceed c, there are other types of velocity that can be greater than c. For example, the phase velocity of a light wave in a medium for which the index of refraction is less than unity is greater than c, but the phase velocity does not correspond to the signal velocity in such a medium; the signal velocity is indeed less than c. Or consider an electron gun that emits a beam of electrons. If the gun is rotated, then the electron beam describes a certain path on a screen placed at some appropriate distance. If the angular velocity of the gun and the distance to the screen are sufficiently large, then the velocity of the spot traveling across the screen can be *any* velocity, arbitrarily large. Thus, the *writing speed* of an oscilloscope can exceed c, but again the writing speed does not correspond to the signal velocity; that is, information cannot be transmitted from one point on the screen to another by means of the electron beam. In such a device, a signal can be transmitted only from the gun to the screen, and this transmission takes place at the velocity of the electrons in the beam (i.e., $< c$).

EXAMPLE 14.11

Derive the relativistic Doppler effect if the angle between the light source and direction of relative motion of the observer is θ (Figure 14-8).

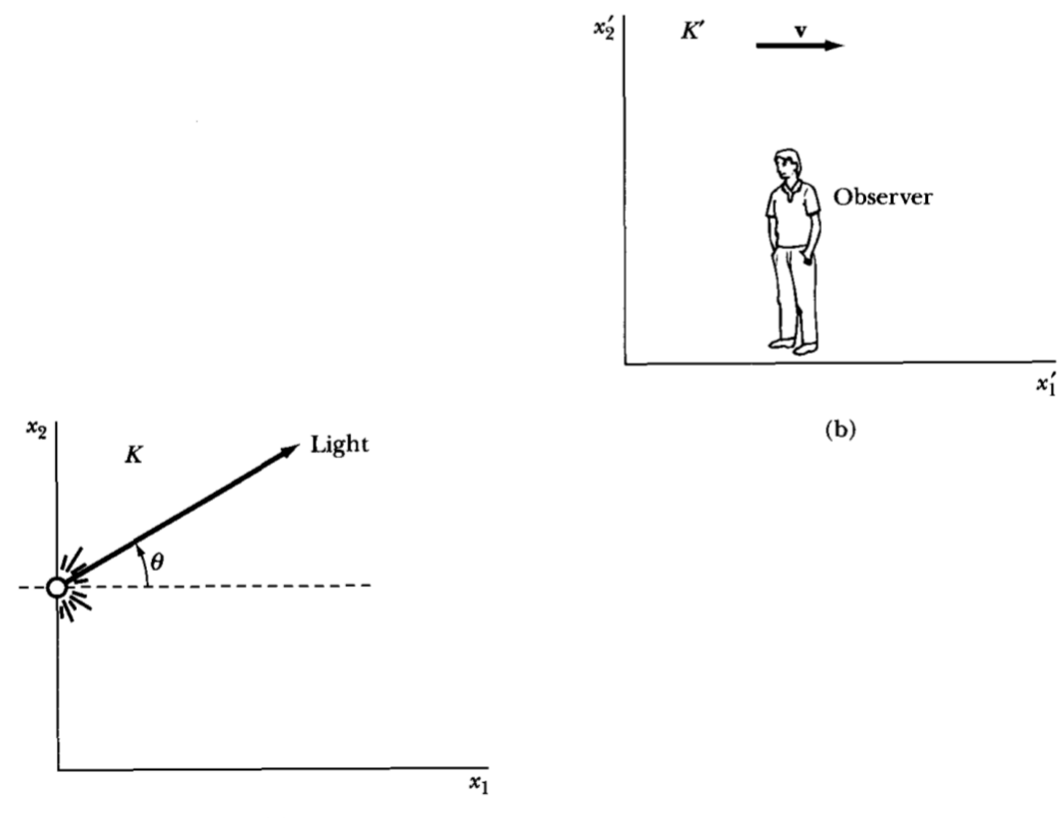


FIGURE 14-8 A light source fixed in system K emits light at a single frequency ν_0. An observer in system K', moving to the right at velocity $\mathbf{v}$ with respect to K, measures the light frequency to be ν'.

Solution. This example can easily be solved using the momentum-energy four-vector by treating the light as a photon with total energy E = $h\nu$. The light source is at rest in system K and emits a single frequency ν_0.

$$E = h\nu_0 \tag{14.99}$$

$$p = \frac{E}{c} = \frac{h\nu_0}{c} \tag{14.100}$$

The observer moving to the right in system K' measures the energy E' for a photon of frequency ν'. From Equation 14.92, we have

$$E' = \gamma(h\nu_0 - vp_1) \tag{14.101}$$

$$h\nu' = \gamma\left(h\nu_0 - \frac{vh\nu_0}{c}\cos\theta\right) \tag{14.102}$$

where $p_1 = p\cos\theta$. Equation 14.102 reduces to

$$\nu' = \gamma\nu_0(1 - \beta\cos\theta) \tag{14.103}$$

which is equivalent to Equation 14.34, depending on the value of θ. For an early time, the observer is far to the left of the source, and as the observer approaches the source ($\theta = \pi$),

$$\nu' = \nu_0\frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \quad \text{observer approaching source} \tag{14.104}$$

as in Equation 14.31. At a much later time, the observer is receding ($\theta = 0$) and

$$\nu' = \nu_0\frac{\sqrt{1-\beta}}{\sqrt{1+\beta}} \quad \text{observer receding from source} \tag{14.105}$$

as in Equation 14.33. When the observer just passes the source ($\theta = \pi/2$),

$$\nu' = \frac{\nu_0}{\sqrt{1-\beta^2}} \qquad \text{observer passing source} \tag{14.106}$$

We can also treat the case where the observer is at rest and the source is moving. We still obtain Equations 14.104 -14.106 because, according to the principle of relativity, it is not possible to distinguish between the motion of the observer and the motion of the source.

14.10 Lagrangian Function in Special Relativity

Lagrangian and Hamiltonian dynamics (discussed in Chapter 7) must be adjusted in light of the new concepts presented here. We can extend the Lagrangian formalism into the realm of special relativity in the following way. For a single (non-relativistic) particle moving in a velocity-independent potential, the rectangular momentum components (see Equation 7.150) may be written as

$$p_i = \frac{\partial L}{\partial u_i} \tag{14.107}$$

According to Equation 14.87, the relativistic expression for the ordinary (i.e., space) momentum component is

$$p_i = \frac{mu_i}{\sqrt{1-\beta^2}} \tag{14.108}$$

We now require that the *relativistic* Lagrangian, when differentiated with respect to u_i as in Equation 14.107, yield the momentum components given by Equation 14.108:

$$\frac{\partial L}{\partial u_i} = \frac{mu_i}{\sqrt{1-\beta^2}} \tag{14.109}$$

This requirement involves only the *velocity* of the particle, so we expect that the velocity-*independent* part of the relativistic Lagrangian is unchanged from the nonrelativistic case. The velocity-*dependent* part, however, may no longer be equal to the kinetic energy. We therefore write

$$L = T^* - U \tag{14.110}$$

where $U = U(x_i)$ and $T^* = T^*(u_i)$. The function T^* must satisfy the relation

$$\frac{\partial T^*}{\partial u_i} = \frac{mu_i}{\sqrt{1-\beta^2}} \tag{14.111}$$

It can be easily verified that a suitable expression for T^* (apart from a possible constant of integration that can be suppressed) is

$$T^* = -mc^2\sqrt{1-\beta^2} \tag{14.112}$$

Hence, the relativistic Lagrangian can be written as

$$\boxed{L = -mc^2\sqrt{1-\beta^2} - U} \tag{14.113}$$

and the equations of motion are obtained in the standard way from Lagrange's equations.

Notice that the Lagrangian is not given by T - U, because the relativistic expression for the kinetic energy (Equation 14.58) is

$$T = \frac{mc^2}{\sqrt{1-\beta^2}} - mc^2 \tag{14.114}$$

The Hamiltonian (see Equation 7.153) can be calculated from

$$\begin{aligned} H &= \sum_i u_i p_i - L \\ &= \sum_i \frac{p_i^2 c^2}{\gamma mc^2} + \frac{mc^2}{\gamma} + U \end{aligned}$$

where we have used Equations 14.108 and 14.113 and changed $\sqrt{1-\beta^2}$ to γ^{-1} . Thus,

$$H = \frac{p^2c^2}{\gamma mc^2} + \frac{mc^2}{\gamma} + U = \frac{1}{\gamma mc^2}(p^2c^2 + m^2c^4) + U$$

$$= \frac{E^2}{\gamma mc^2} + U$$

$$= E + U = T + U + E_0 \tag{14.115}$$

The relativistic Hamiltonian is equal to the total energy defined in Section 14.8 plus the potential energy. It differs from the total energy used previously in Chapter 7 by now including the rest energy.

14.11 Relativistic Kinematics

In the event that the velocities in a collision process are not negligible with respect to the velocity of light, it becomes necessary to use relativistic kinematics. In the discussion in Chapter 9, we took advantage of the properties of the center-of-mass coordinate system in deriving many of the kinematic relations. Because mass and energy are interrelated in relativity theory, it no longer is meaningful to speak of a "center-of-mass" system; in relativistic kinematics, one uses a "center-of-momentum" coordinate system instead. Such a system possesses the same essential property as the previously used center-of-mass system—the total linear momentum in the system is zero. Therefore, if a particle of mass m_1 collides elastically with a particle of mass m^, then in the center-of-momentum system we have

$$p_1' = p_2' \tag{14.116}$$

Using Equation 14.87, the space components of the momentum four-vector can be written as

$$m_1 u_1' \gamma_1' = m_2 u_2' \gamma_2' \tag{14.117}$$

where, as before, $\gamma = 1/\sqrt{1-\beta^2}$ and $\beta = u/c$.

In a collision problem, it is convenient to associate the laboratory coordinate system with the inertial system K and the center-of-momentum system with K' (see Figure 14-9).

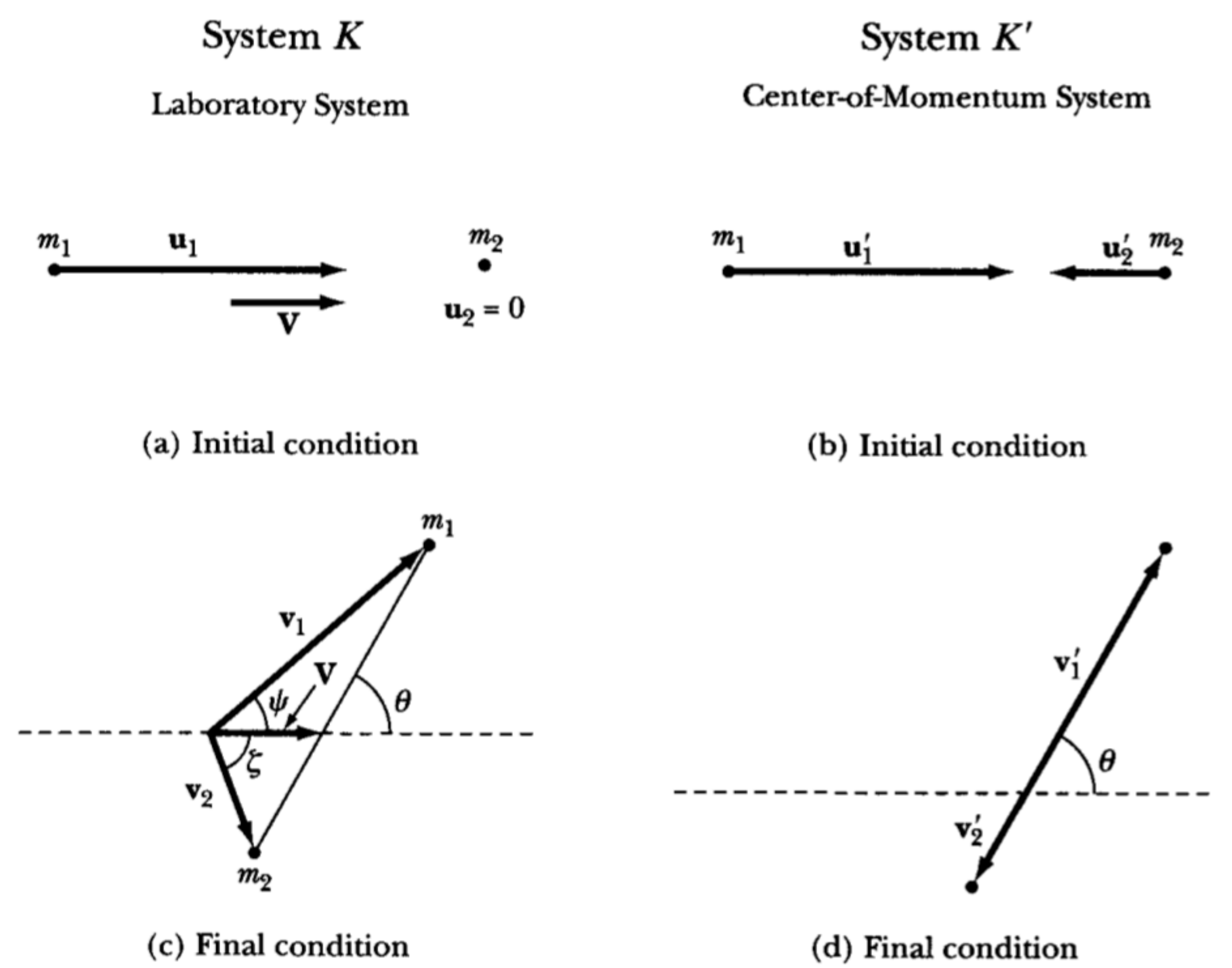


FIGURE 14-9 The elastic collision schematic of Figure 9-10 is redisplayed with systems K and K' indicated.

A simple Lorentz transformation then connects the two systems. To derive the relativistic kinematic expressions, the procedure is to obtain the center-of-momentum relations and then perform a Lorentz transformation back to the laboratory system. We choose the coordinate axes so that m_1 moves along the x-axis in K with speed u_1. Because m_2 is initially at rest in K, $u_2 = 0$. In K', m_2 moves with speed u'_2 and so K' moves with respect to K also with speed u'_2 and in the same direction as the initial motion of m_1.

Using the fact that $\beta\gamma = \sqrt{\gamma^2 - 1}$, we have

$$\begin{aligned} p'_1 &= m_1 u'_1 \gamma'_1 = m_1 c \beta'_1 \gamma'_1 \\ &= m_1 c \sqrt{\gamma_1'^2 - 1} = m_2 c \sqrt{\gamma_2'^2 - 1} \\ &= p'_2 \end{aligned} \tag{14.118}$$

which expresses the equality of the momenta in the center-of-momentum system.

According to Equation 14.92, the transformation of the momentum p_i (from K to K') is

$$p_1' = \left(p_1 - \frac{u_2'}{c^2} E_1\right)\gamma_2' \tag{14.119}$$

We also have

$$\left.\begin{aligned} p_1 &= m_1 u_1 \gamma_1 \\ E_1 &= m_1 c^2 \gamma_1 \end{aligned}\right\} \tag{14.120}$$

so Equation 14.118 can be used to obtain

$$\begin{aligned} m_1 c \sqrt{\gamma_1'^2 - 1} &= (m_1 c \beta_1 \gamma_1 - \beta_2' m_1 c \gamma_1)\gamma_2' \\ &= m_1 c\left(\gamma_2' \sqrt{\gamma_1^2 - 1} - \gamma_1 \sqrt{\gamma_2'^2 - 1}\right) \\ &= m_2 c \sqrt{\gamma_2'^2 - 1} \end{aligned} \tag{14.121}$$

These equations can be solved for γ_1' and γ_2' in terms of γ_1 :

$$\gamma_1' = \frac{\gamma_1 + \dfrac{m_1}{m_2}}{\sqrt{1 + 2\gamma_1\left(\dfrac{m_1}{m_2}\right) + \left(\dfrac{m_1}{m_2}\right)^2}} \tag{14.122a}$$

$$\gamma_2' = \frac{\gamma_1 + \dfrac{m_2}{m_1}}{\sqrt{1 + 2\gamma_1\left(\dfrac{m_2}{m_1}\right) + \left(\dfrac{m_2}{m_1}\right)^2}} \tag{14.122b}$$

Next, we write the equations of the transformation of the momentum components from K' back to K after the scattering. We now have both x- and y-components:

$$p_{1,x} = \left(p'_{1,x} + \frac{u'_2}{c^2} E'_1\right)\gamma'_2$$

$$= (m_1 c \beta'_1 \gamma'_1 \cos\theta + m_1 c \beta'_2 \gamma'_1)\gamma'_2 \tag{14.123a}$$

$$= m_1 c \gamma'_1 \gamma'_2 (\beta'_1 \cos\theta + \beta'_2)$$

(Note that, because the transformation is from K' to K, a plus sign occurs before the second term, in contrast to Equation 14.119.) Also,

$$p_{1,y} = m_1 c \beta'_1 \gamma'_1 \sin\theta \tag{14.123b}$$

The tangent of the laboratory scattering angle ψ is given by $p_{i,y}/p_{i,x}$, therefore, dividing Equation 14.123b by Equation 14.123a, we obtain

$$\tan\psi = \frac{1}{\gamma'_2} \frac{\sin\theta}{\cos\theta + (\beta'_2/\beta'_1)}$$

Using Equation 14.117 to express β'_2/β'_1, the result is

$$\tan\psi = \frac{1}{\gamma'_2} \frac{\sin\theta}{\cos\theta + (m_1\gamma'_1/m_2\gamma'_2)} \tag{14.124}$$

For the recoil particle, we have

$$p_{2,x} = \left(p'_{2,x} + \frac{u'_2}{c^2} E'_2\right)\gamma'_2$$

$$= (-m_2 c \beta'_2 \gamma'_2 \cos\theta + m_2 c \beta'_2 \gamma'_2)\gamma'_2 \tag{14.125a}$$

$$= m_2 c \beta'_2 \gamma'^2_2 (1 - \cos\theta)$$

where a minus sign occurs in the first term because $p'_{2,x}$ is directed opposite to $p_{1,x}$. Also,

$$p_{2,y} = -m_2 c \beta_2' \gamma_2' \sin\theta \tag{14.125b}$$

As before, the tangent of the laboratory recoil angle ζ is given by $p_{2,y}/p_{2,x}$:

$$\tan\zeta = -\frac{1}{\gamma_2'}\frac{\sin\theta}{1 - \cos\theta} \tag{14.126}$$

The overall minus sign indicates that if m_1 is scattered toward positive values of ψ , then m_2 recoils in the negative ζ-direction.

A case of special interest is that in which $m_1 = m_2$. From Equations 14.122, we find

$$\gamma_1' = \gamma_2' = \sqrt{\frac{1 + \gamma_1}{2}}, \qquad m_1 = m_2 \tag{14.127}$$

The tangents of the scattering angles become

$$\tan\psi = \sqrt{\frac{2}{1 + \gamma_1}} \cdot \frac{\sin\theta}{1 + \cos\theta} \tag{14.128}$$

$$\tan\zeta = -\sqrt{\frac{2}{1 + \gamma_1}} \cdot \frac{\sin\theta}{1 - \cos\theta} \tag{14.129}$$

The product is therefore

$$\tan\psi\tan\zeta = -\frac{2}{1 + \gamma_1}, \qquad m_1 = m_2 \tag{14.130}$$

(The minus sign is of no essential importance; it only indicates that ψ and ζ are measured in opposite directions.)

We previously found that in the nonrelativistic limit there was always a right angle between the final velocity vectors in the scattering of particles of equal mass. Indeed, in the limit $\gamma_1 \to 1$, Equations 14.128 and 14.129 become equal to Equations 9.69 and 9.73, respectively, and so $\psi + \zeta = \pi/2$. Equation 14.130, however, shows that in the relativistic case $\psi + \zeta < \pi/2$; thus, the included angle in the scattering is always smaller than in the nonrelativistic limit. For equal scattering and recoil angles ($\psi = \zeta$), Equation 14.130 becomes

$$\tan\psi = \left(\frac{2}{1+\gamma_1}\right)^{1/2}, \qquad m_1 = m_2$$

and the included angle between the directions of the scattered and recoil particles is

$$\begin{aligned} \phi &= \psi + \zeta = 2\psi \\ &= 2\tan^{-1}\left(\frac{2}{1+\gamma_1}\right)^{1/2}, \qquad m_1 = m_2 \end{aligned} \tag{14.131}$$

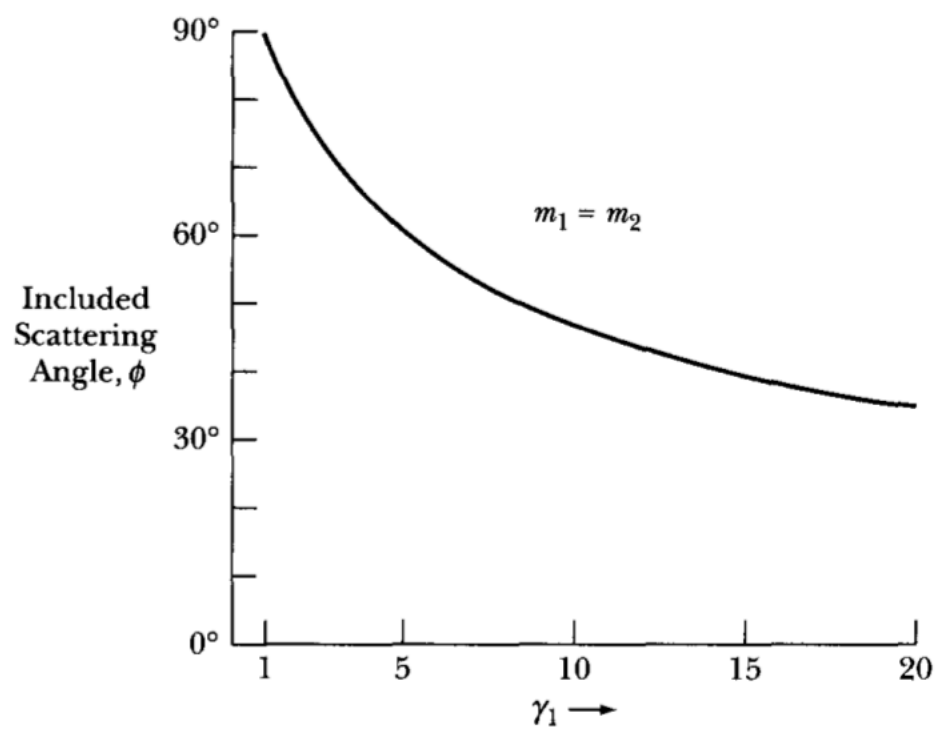


FIGURE 14-10 The included scattering angle, $\phi = \psi + \zeta$, is shown as a function of the relativistic parameter γ_1 for $m_1 = m_2$. For nonrelativistic scattering ($\gamma_1 = 1$), this angle is always 90°.

Figure 14-10 shows ϕ as a function of γ_1 up to $\gamma_1 = 20$. At $\gamma_1 = 10$, the included angle is approximately 46°. This value of γ_1 corresponds to an initial velocity that is 99.5% of the velocity of light. According to Equation 14.58, the kinetic energy is given by $T_1 = m_1c^2(\gamma_1 - 1)$; therefore, a proton with $\gamma_1 = 10$ would have a kinetic energy of approximately 8.4 GeV, whereas an electron with the same velocity would have $T_1 = 4.6$ MeV.

By using the transformation properties of the fourth component of the momentum four-vector (i.e., the total energy), it is possible to obtain the relativistic analogs of all the energy equations we have previously derived in the nonrelativistic limit.