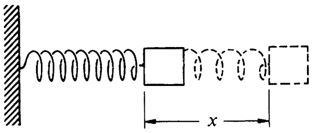
10 THE HARMONIC OSCILLATOR

10.1 Introduction and Review

The motion of a mass on a spring, better known as a harmonic oscillator, is familiar to us from Chaps. 2 and 4 and from numerous problems. However, so far we have considered only the idealized case in which friction is absent and there are no external forces. In this chapter we shall investigate the effect of friction on the oscillator, and then study the motion when the mass is subjected to a driving force which is a periodic function of time. Finally, we shall use the harmonic oscillator to illustrate a remarkable result—the possibility of predicting how a mechanical system will respond to an applied driving force of any given frequency merely by studying what the system does when it is put into motion and allowed to move freely.



We begin by reviewing the properties of the frictionless harmonic oscillator which we discussed at the end of Chap. 2. The prototype oscillator is a mass m acted on by a spring force $F_{spring} = -kx$, where x is the displacement from equilibrium. The equation of motion is $m\ddot{x} = -kx$, or

$$m\ddot{x} + kx = 0. \qquad 10.1$$

The solution is

$$x = B \sin \omega_0 t + C \cos \omega_0 t, \qquad 10.2$$

where

$$\omega_0 = \sqrt{\frac{k}{m}}. \qquad 10.3$$

We shall use ω_0 rather than ω, as in previous chapters, to represent the natural frequency of the oscillator. B and C are arbitrary constants which can be evaluated from a set of given initial conditions, such as the position and the velocity at a particular time.

Standard Form of the Solution

We can rewrite Eq. (10.2) in the following more convenient form:

$$x = A \cos (\omega_0 t + \phi), \qquad 10.4$$

where A and ϕ are constants. To show the correspondence between Eqs. (10.2) and (10.4) we make use of the trigonometric identity

$$\cos (\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta.$$

By applying this to Eq. (10.4) and equating Eqs. (10.2) and (10.4), we obtain

$$A \cos \omega_0 t \cos \phi - A \sin \omega_0 t \sin \phi = B \sin \omega_0 t + C \cos \omega_0 t.$$

For this to be true at all times, the coefficients of the terms in $\sin \omega_0 t$ and $\cos \omega_0 t$ must be separately equal. Hence, we have

$$A \cos \phi = C$$

$$A \sin \phi = -B, \qquad 10.5a$$

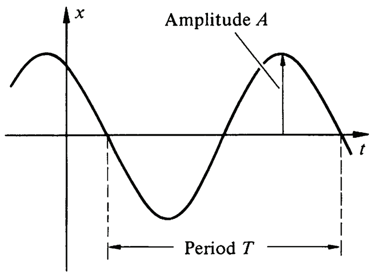
which are readily solved to yield

$$A = (B^2 + C^2)^{\frac{1}{2}}$$

$$\tan \phi = -\frac{B}{C}. \qquad 10.5b$$

This result shows that the two expressions Eqs. (10.2) and (10.4) for the general motion of the harmonic oscillator are equivalent. We shall generally use Eq. (10.4) as the standard form for the motion of a frictionless harmonic oscillator.

Nomenclature



There are a number of definitions with which we should be familiar. Consider the expression

$$x = A \cos (\omega_0 t + \phi).$$

x is the instantaneous displacement of the particle at time t.

A is the amplitude of the motion, measured from zero displacement to a maximum.

ω_0 is the frequency (or angular frequency) of motion. $\omega_0 = \sqrt{k/m}$ rad/s. The circular frequency $\nu = \omega_0/2\pi$ Hz (1 Hz = 1 cycle per second).

ϕ is the phase factor or phase angle.

T is the period of the motion, the time required to execute one complete cycle. $T = 2\pi/\omega_0$.

Example 10.1 **Initial Conditions and the Frictionless Harmonic Oscillator**

Suppose that at time $t = 0$ the position of the mass is $x(0)$ and its velocity $v(0)$. If we express the motion in the form of Eq. (10.2) we have

$$x = B \sin \omega_0 t + C \cos \omega_0 t$$

$$v = \dot{x}$$

$$= \omega_0 B \cos \omega_0 t - \omega_0 C \sin \omega_0 t.$$

Evaluating these at $t = 0$ gives

$$C = x(0)$$

$$B = \frac{\dot{x}(0)}{\omega_0}.$$

If we begin with the standard form $x = A \cos(\omega_0 t + \phi)$, the displacement and velocity are

$$x = A \cos(\omega_0 t + \phi)$$

$$v = -\omega_0 A \sin(\omega_0 t + \phi).$$

For $t = 0$,

$$x(0) = A \cos \phi$$

$$v(0) = -\omega_0 A \sin \phi,$$

from which we find

$$A = \sqrt{x(0)^2 + \left[\frac{v(0)}{\omega_0}\right]^2}$$

$$\tan \phi = \frac{-v(0)}{\omega_0 x(0)}.$$

Energy Considerations

If we take the potential energy to be 0 at $x = 0$, we have

$$U = \tfrac{1}{2}kx^2$$

$$= \tfrac{1}{2}kA^2 \cos^2(\omega_0 t + \phi). \qquad 10.6a$$

The kinetic energy is

$$K = \tfrac{1}{2}mv^2$$

$$= \tfrac{1}{2}m\omega_0^2 A^2 \sin^2(\omega_0 t + \phi), \qquad 10.6b$$

where we have used

$$v = \dot{x} = -\omega_0 A \sin(\omega_0 t + \phi).$$

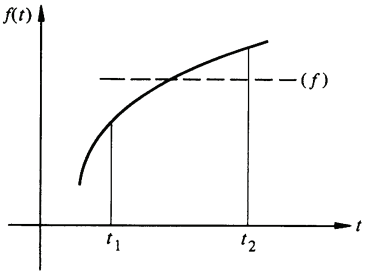
Since $\omega_0^2 = k/m$, Eq. (10.6b) becomes $K = \frac{1}{2}kA^2 \sin^2(\omega_0 t + \phi)$. The total energy is

$$E = K + U = \tfrac{1}{2}kA^2 [\cos^2(\omega_0 t + \phi) + \sin^2(\omega_0 t + \phi)]$$

$$E = \tfrac{1}{2}kA^2. \qquad 10.7$$

Hence, the total energy is constant, a familiar feature of motion when only conservative forces act.

Time Average Values



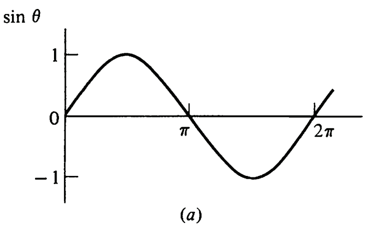
In the following sections we need the concept of the *time average value* $\langle f \rangle$ of a function $f(t)$. Consider $f(t)$, some function of time, and an interval $t_1 \le t \le t_2$ as shown in the sketch. $\langle f \rangle$, the time average value of $f(t)$, is defined so that the rectangular area shown in the sketch, $(t_2 - t_1)\langle f \rangle$, equals the actual area under the curve between t_1 and t_2:

$$(t_2 - t_1)\langle f \rangle = \int_{t_1}^{t_2} f(t)\, dt$$

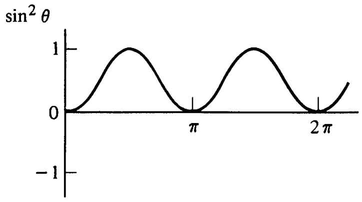
or

$$\langle f \rangle = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} f(t)\, dt.$$

To make this idea more concrete, suppose that $f(t)$ represents the rate of flow of water into a bucket in liters per second. Then the volume of water passing into the bucket in a short interval dt is $f(t)\, dt$, and the total volume passing into the bucket in the interval $t_2 - t_1$ is $\int_{t_1}^{t_2} f(t)\, dt$. If the flow were steady, the rate would have to be $\langle f \rangle$ for the same volume of water to accumulate in the time interval $t_2 - t_1$.



(a)



(b)

For our work with the harmonic oscillator we shall need the time averages of sin (ωt) and $\sin^2 (\omega t)$ over one cycle of oscillation. Here is a graphical device for calculating these averages. The first sketch shows $\sin \theta$ for the interval $0 \le \theta \le 2\pi$, where $\theta = \omega t$. It is apparent that the area above the axis equals the area below the axis, so that $\langle \sin \theta \rangle = 0$. In the second sketch, we show $\sin^2 \theta$. This varies between 0 and 1, and by symmetry we see that its average value is $\frac{1}{2}$. Thus $\langle \sin^2 \theta \rangle = \frac{1}{2}$. By identical arguments $\langle \cos \theta \rangle = 0$, $\langle \cos^2 \theta \rangle = \frac{1}{2}$, and you can also show graphically that these results hold as long as the average is taken over a whole period of oscillation, irrespective of the starting point. These results can also be proven analytically; we leave this task for a problem.

Average Energy

Returning to the frictionless harmonic oscillator, we can now evaluate the time average values of the potential and kinetic energies over one period of oscillation $0 \leq t \leq T$. From Eq. (10.6a),

$$U = \tfrac{1}{2}kA^2 \cos^2(\omega_0 t + \phi)$$
$$\langle U\rangle = \tfrac{1}{2}kA^2 \langle\cos^2(\omega_0 t + \phi)\rangle$$
$$= \tfrac{1}{4}kA^2.$$

(We have used $\langle\cos^2\theta\rangle = \frac{1}{2}$ for an average over one period.) Similarly, from Eq. (10.6b),

$$\langle K\rangle = \tfrac{1}{2}m\omega_0{}^2A^2 \langle\sin^2(\omega_0 t + \phi)\rangle$$
$$= \tfrac{1}{4}m\omega_0{}^2A^2.$$

Since $\omega_0{}^2 = k/m$, we have

$$\langle K\rangle = \tfrac{1}{4}kA^2$$
$$= \langle U\rangle.$$

The time average kinetic and potential energies are equal. When friction is present, this is no longer exactly true.

10.2 The Damped Harmonic Oscillator

Our next step is to consider the effect of friction on the harmonic oscillator. We are going to restrict our discussion to a very special form of friction force, the viscous force. Such a force arises when an object moves through a fluid, either liquid or gas, at speeds which are not so large as to cause turbulence. In this case the friction force f is of the form

$$f = -bv,$$

where b is a constant of proportionality that depends on the shape of the mass and the medium through which it moves, and where v is the instantaneous velocity. Although this is a special friction force, we should emphasize that it is the type most often encountered and that our analysis has wide applicability. Although the discussion here is devoted to a mechanical oscillator, equations of identical form describe many other oscillating systems. For example, electric current can oscillate in certain electric circuits; the electrical resistance of the circuit plays a role exactly analogous to a viscous retarding force.

The total force acting on the mass m is

$$F = F_{\text{spring}} + f$$
$$= -kx - bv.$$

The equation of motion is

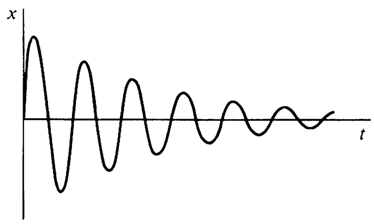
$$m\ddot{x} = -kx - b\dot{x},$$

which can be rewritten as

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0. \qquad 10.8$$

Here γ stands for b/m and, as before, $\omega_0^2 = k/m$. The units of γ are second^{-1}.

Equation (10.8) is a more complicated differential equation than any we have yet encountered. We leave the details of the solution for Note 10.1 and merely state the result here:



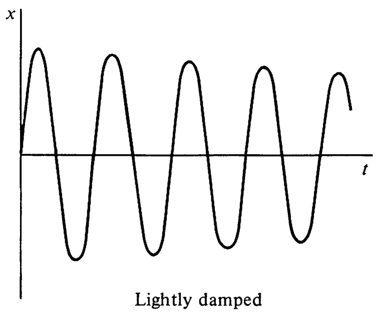
$$x = Ae^{-(\gamma/2)t}\cos(\omega_1 t + \phi). \qquad 10.9$$

A and ϕ again stand for arbitrary constants and

$$\omega_1 = \sqrt{\omega_0^2 - \frac{\gamma^2}{4}}. \qquad 10.10$$

This solution is valid when $\omega_0^2 - \gamma^2/4 > 0$, or, equivalently, $\gamma < 2\omega_0$. (Other cases are discussed in Note 10.1). Substituting Eq. 10.9 into Eq. (10.8) to verify the solution makes a good exercise.

The motion described by Eq. (10.9) is known as *damped harmonic motion*. A typical case is shown in the top sketch. The motion is reminiscent of the undamped harmonic motion described in the last section. In fact, we can rewrite Eq. (10.9) as



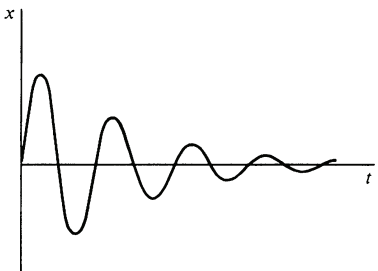
Lightly damped

$$x = A(t)\cos(\omega_1 t + \phi),$$

where

$$A(t) = Ae^{-(\gamma/2)t}. \qquad 10.11$$

The motion is similar to the undamped case except that the amplitude decreases exponentially in time and the frequency of oscillation ω_1 is less than the undamped frequency ω_0. Incidentally, although the concept of a definite frequency can be strictly applied only to a pure sine or cosine function, ω_1 is commonly called the frequency of oscillation. The zero crossings of the function $Ae^{-(\gamma/2)t}\cos(\omega_1 t + \phi)$ are separated by equal time intervals $T = 2\pi/\omega_1$, but the peaks do not lie halfway between them.



Heavily damped

Before we investigate damped harmonic motion quantitatively, it will be helpful to look at it qualitatively. The essential features of the motion depend on the ratio γ/ω_1. If $\gamma/\omega_1 \ll 1$, $A(t)$ decreases very little during the time the cosine makes many zero crossings; in this regime, the motion is called lightly damped.

If γ/ω_1 is comparatively large, $A(t)$ tends rapidly to zero while the cosine makes only a few oscillations. This motion is called heavily damped. For light damping, $\omega_1 \approx \omega_0$, but for heavy damping ω_1 can be significantly smaller than ω_0.

Energy

By considering the energy of the system we can see why the amplitude must decrease with time. From the work-energy theorem of Chap. 4,

$$E(t) = E(0) + W_{\text{frict}},$$

where

$$E(t) = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = K(t) + U(t)$$

and

$$W_{\text{frict}} = \text{work done by friction from time 0 to time } t.$$

The dissipative friction force, $f = -bv$, opposes the motion. Hence,

$$\begin{aligned} W_{\text{frict}} &= \int_{x(0)}^{x(t)} f\,dx \\ &= \int_0^t fv\,dt \\ &= -\int_0^t bv^2\,dt < 0. \end{aligned} \qquad 10.12$$

Physically, $E(t)$ decreases with time because the friction force continually dissipates energy. We can find how $E(t)$ depends on time by calculating the kinetic and potential energies $K(t)$ and $U(t)$.

To evaluate $K(t) = \frac{1}{2}mv^2$ we need the velocity v. The time derivative of Eq. (10.9) gives

$$\begin{aligned} v &= -Ae^{-(\gamma/2)t}\left[\omega_1 \sin(\omega_1 t + \phi) + \frac{\gamma}{2}\cos(\omega_1 t + \phi)\right] \\ &= -\omega_1 A e^{-(\gamma/2)t}\left[\sin(\omega_1 t + \phi) + \frac{1}{2}\left(\frac{\gamma}{\omega_1}\right)\cos(\omega_1 t + \phi)\right]. \end{aligned} \qquad 10.13$$

If the motion is only lightly damped, $\gamma/\omega_1 \ll 1$, and the coefficient of the second term in the bracket is small. Let us assume that the damping is so small that we can neglect the second term entirely. In this case we have

$$v = -\omega_1 A e^{-(\gamma/2)t} \sin(\omega_1 t + \phi),$$

and

$$K(t) = \tfrac{1}{2}mv^2$$
$$= \tfrac{1}{2}m\omega_1^2 A^2 e^{-\gamma t} \sin^2(\omega_1 t + \phi). \qquad 10.14a$$

The potential energy is

$$U(t) = \tfrac{1}{2}kx^2$$
$$= \tfrac{1}{2}kA^2 e^{-\gamma t} \cos^2(\omega_1 t + \phi) \qquad 10.14b$$

and the total energy is

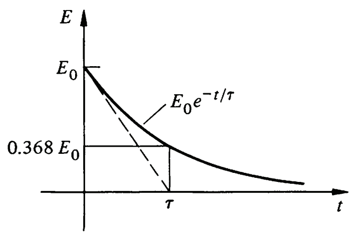
$$E(t) = K(t) + U(t)$$
$$= \tfrac{1}{2}A^2 e^{-\gamma t}[m\omega_1^2 \sin^2(\omega_1 t + \phi) + k\cos^2(\omega_1 t + \phi)].$$

Since the damping is assumed to be small, we can simplify the term in brackets by replacing ω_1^2 by ω_0^2† and using the relation $\omega_0^2 = k/m$.

$$E(t) = \tfrac{1}{2}A^2 e^{-\gamma t}[k\cos^2(\omega_1 t + \phi) + k\sin^2(\omega_1 t + \phi)]$$
$$= \tfrac{1}{2}kA^2 e^{-\gamma t} \qquad 10.15$$

At $t = 0$ the energy of the system is

$$E_0 = \tfrac{1}{2}kA^2$$



and we can rewrite Eq. (10.15) as

$$E(t) = E_0 e^{-\gamma t}. \qquad 10.16$$

This is a remarkably simple result. The energy decreases exponentially in time.

The decay can be characterized by the time τ required for the energy to drop to $e^{-1} = 0.368$ of its initial value.

$$E(\tau) = E_0 e^{-\gamma\tau}$$
$$= e^{-1}E_0.$$

† This approximation can be justified for $\gamma/\omega_1 \ll 1$ as follows:

$$\omega_0^2 = \omega_1^2 + \frac{\gamma^2}{4}$$
$$= \omega_1^2\left[1 + \frac{1}{4}\left(\frac{\gamma}{\omega_1}\right)^2\right]$$
$$\approx \omega_1^2.$$

Hence, $\gamma\tau = 1$.

$$\tau = \frac{1}{\gamma} = \frac{m}{b}. \qquad 10.17$$

τ is often called the *damping time* (or, alternatively, the *time constant* or characteristic time) of the system. In the limit of light damping, $\gamma \to 0$ and $\tau \to \infty$; E is effectively constant and the system behaves like an undamped oscillator.

The Q of an Oscillator

The degree of damping of an oscillator is often specified by a dimensionless parameter Q, the *quality factor*, defined by

$$Q = \frac{\text{energy stored in the oscillator}}{\text{energy dissipated per radian}}. \qquad 10.18$$

By energy dissipated per radian we mean the energy lost during the time it takes the system to oscillate through one radian. In the period $T = 2\pi/\omega_1$, the system oscillates through 2π radians. Thus the time to oscillate through one radian is $T/2\pi = 1/\omega_1$.

Q is easily calculated for the lightly damped case. The rate of change of energy is, from Eq. (10.16),

$$\begin{aligned}\frac{dE}{dt} &= -\gamma E_0 e^{-\gamma t}\\ &= -\gamma E.\end{aligned}$$

The energy dissipated in a short time Δt is the positive quantity

$$\begin{aligned}\Delta E &\approx \left|\frac{dE}{dt}\right| \Delta t\\ &= \gamma E\, \Delta t.\end{aligned}$$

One radian of oscillation requires time $\Delta t = 1/\omega_1$, and the energy dissipated is $\gamma E/\omega_1$. Hence, the quality factor is

$$Q = \frac{E}{\gamma E/\omega_1} = \frac{\omega_1}{\gamma} \approx \frac{\omega_0}{\gamma}. \qquad 10.19$$

A lightly damped oscillator has $Q \gg 1$. A heavily damped system loses its energy rapidly and its Q is low. A tuning fork has a Q of a thousand or so, whereas a superconducting microwave cavity can have a Q in excess of 10^7. An undamped oscillator has infinite Q.

Example 10.2 The Q of Two Simple Oscillators

A musician's tuning fork rings at A above middle C, 440 Hz. A sound level meter indicates that the sound intensity decreases by a factor of 5 in 4 s. What is the Q of the tuning fork?

The sound intensity from the tuning fork is proportional to the energy of oscillation. Since the energy of a damped oscillator decreases as $e^{-\gamma t}$, we can find γ by taking the ratio of the energy at $t = 0$ to that at $t = 4$ s.

$$5 = \frac{E(0)e^{(0)}}{E(0)e^{-4\gamma}} = e^{4\gamma}$$

Hence

$$4\gamma = \ln 5 = 1.6$$
$$\gamma = 0.4 \text{ s}^{-1},$$

and

$$Q = \frac{\omega_1}{\gamma} = \frac{2\pi(440)}{0.4}$$
$$\approx 700.$$

The energy loss is due primarily to the heating of the metal as it bends. Air friction and energy loss to the mounting point also contribute. (The symmetrical design of a tuning fork minimizes loss to the mount.) Incidentally, if you try this experiment, bear in mind that the ear is a poor sound level meter because it does not respond linearly to sound intensity; its response is more nearly logarithmic.

A rubber band exhibits a much lower Q than a tuning fork primarily because of the internal friction generated by the coiling of the long chain molecules. In one experiment, a paperweight suspended from a hefty rubber band had a period of 1.2 s and the amplitude of oscillation decreased by a factor of 2 after three periods. What is the estimated Q of this system?

From Eq. (10.11) the amplitude is given by $Ae^{(-\gamma/2)t}$. The ratio of the amplitude at $t = 0$ to that at $t = 3(1.2) = 3.6$ s is

$$2 = \frac{Ae^{(0)}}{Ae^{-3.6\gamma/2}}.$$

Hence

$$1.8\gamma = \ln 2 = 0.69$$

or

$$\gamma = 0.39 \text{ s}^{-1}.$$

Therefore

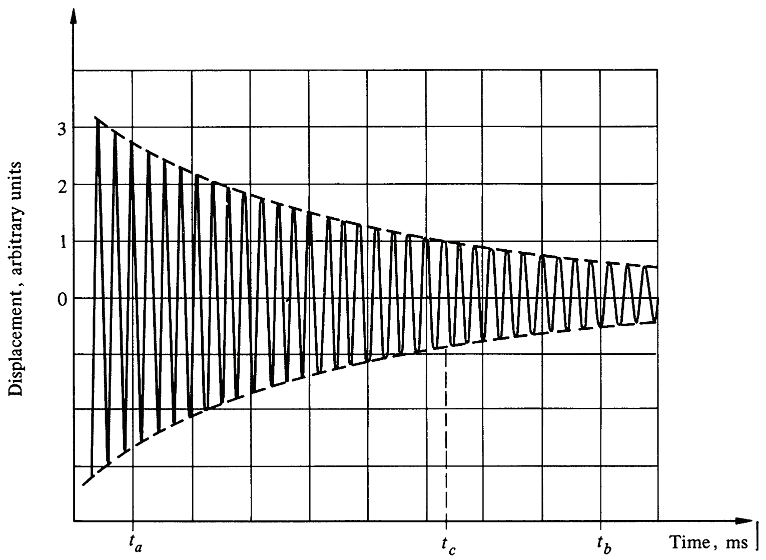
$$Q \approx \frac{\omega_1}{\gamma}$$
$$= \frac{2\pi/T}{0.39}$$
$$= \frac{2\pi/1.2}{0.39}$$
$$= 13.$$

You may wonder whether it is justifiable to use the light damping result, $Q = \omega_1/\gamma$, when Q is so low. The approximations involved introduce errors of order $(\gamma/\omega_1)^2 = (1/Q)^2$. For $Q > 10$ the error is less than 1 percent.

It is interesting to note that the damping constants for the tuning fork and for the rubber band are very nearly the same. The tuning fork has a much higher Q, however, because it goes through many more cycles of oscillation in one damping time and loses correspondingly less of its energy per cycle.

Example 10.3 **Graphical Analysis of a Damped Oscillator**

The illustration is drawn from a photograph of an oscilloscope trace of the displacement of an oscillating system versus time. We immediately recognize that the system is a damped harmonic oscillator. The frequency ω_1 and quality factor Q can be found from the photograph.



The time interval from t_a to t_b is 8 ms. There are 28.5 cycles (i.e., complete periods) in this interval. (Check this for yourself from the data.) The period of oscillation is $T = 8 \times 10^{-3}$ s/28.5 = 2.81×10^{-4} s. The angular frequency is $\omega_1 = 2\pi/T = 22{,}400$ rad/s. The corresponding circular frequency is $\nu = \omega_1/2\pi = 3{,}560$ Hz.

In order to obtain the quality factor $Q = \omega_1/\gamma$, the damping constant must be known. From Eq. (10.11) the amplitude is $Ae^{-(\gamma/2)t}$. This function describes the *envelope* of the displacement curve. The envelope has been drawn with a dashed curve on the photograph. At time t_a the envelope has magnitude $A_a = 2.75$ units. When the envelope decays by a factor $e^{-1} = 0.368$, its magnitude is 1.01 units. From the photograph this occurs at $t_c = 5.35$ ms, measured from t_a. Hence, $e^{-(\gamma/2)t_c} = e^{-1}$, or $\gamma = 2/t_c = 374\ \text{s}^{-1}$. The quality factor is $Q = \omega_1/\gamma = 60$.

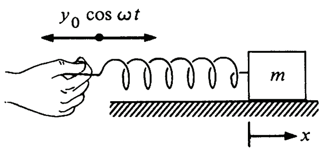
Now for a word about the system. This is not a mechanical oscillator, nor even an electrical oscillator. The signal is produced by radiating atomic electrons in a small volume of hydrogen gas. The signal was greatly amplified for oscilloscope display. Furthermore, the atoms were actually radiating at 9.2×10^9 Hz. Since this is much too high for the oscilloscope to follow, the frequency was translated to a lower value by electronic means. This did not affect the shape of the envelope, and our measured value of γ is correct. If we use the true value of the frequency of the atomic system, we find that the actual Q is

$$Q = \frac{2\pi\nu}{\gamma}$$
$$= \frac{2\pi \times 9.2 \times 10^9}{374}$$
$$= 1.6 \times 10^8.$$

Such a high Q is virtually unattainable for mechanical systems, although it is not unusual in an atomic system.

10.3 The Forced Harmonic Oscillator

The Undamped Forced Oscillator



We next investigate the effect of an applied time varying force $F(t)$ on a frictionless harmonic oscillator. In the case of a mass on a spring, the force can be applied by jiggling the end of the spring. To be concrete, suppose that the end of the spring moves according to $y = y_0 \cos \omega t$, as shown in the sketch. The change in length of the spring from its equilibrium length is $x - y$, where x is the position of the mass. The equation of motion, neglecting damping, is $m\ddot{x} = -k(x - y)$, or, since $y = y_0 \cos \omega t$,

$$m\ddot{x} + kx = F_0 \cos \omega t \qquad 10.20$$

where $F_0 = ky_0$. $F_0 \cos \omega t$ is called the *driving force*. F_0 is the amplitude of the driving force (note that F_0 has the dimensions of force) and ω is the driving frequency, a quantity we are free to vary.

It is apparent that we have chosen a very special form for the driving force in Eq. (10.21). Nevertheless, the solution is of quite general interest. It turns out that any periodic function of time can be represented as a sum of sinusoidal terms (this is the basis of Fourier's theorem), so that understanding the response of the harmonic oscillator to the force $F_0 \cos \omega t$ lays the groundwork for finding the response to any periodic force. Furthermore, many important cases involve the simple sinusoidal force we assume here; two examples are the response of a bound electron to an electromagnetic field (a problem which arises in the classical theory of the scattering of light) and the tidal response of a lake to the periodic force of the moon or sun. So, without further apology we turn to the solution of Eq. (10.20).

A general procedure for solving Eq. (10.20) is given in Note 10.2, but in fact this equation is so simple that we can guess the correct solution by the following argument: the right hand side of the equation varies in time as $\cos \omega t$. It seems plausible that the left hand side involves the same time dependence. We try the solution

$$x = A \cos \omega t.$$

Substituting this in Eq. (10.20) yields

$$(-m\omega^2 + k)A \cos \omega t = F_0 \cos \omega t,$$

which is valid provided that we choose

$$A = \frac{F_0}{k - m\omega^2} = \frac{F_0}{m}\frac{1}{\omega_0^2 - \omega^2}, \qquad 10.21$$

where $\omega_0^2 = k/m$, as in the last section. Our solution becomes

$$x = \frac{F_0}{m}\frac{1}{\omega_0^2 - \omega^2}\cos \omega t. \qquad 10.22$$

The solution we found in Eq. (10.22) is quite different in nature from the solution of Eq. (10.4) or (10.9). There are no arbitrary constants in Eq. (10.22); the motion is fully determined. Physically, this is surprising, since we should be able to specify the initial position and velocity of any particle obeying Newton's laws. The difficulty is that although the solution in Eq. (10.22) is correct, it is not complete. The complete solution is[1]

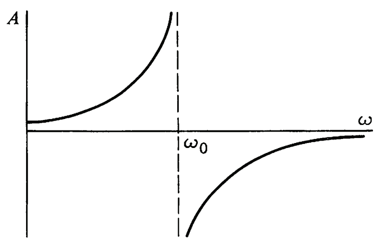
$$x = \frac{F_0}{m}\frac{1}{\omega_0^2 - \omega^2}\cos\omega t + B\cos(\omega_0 t + \phi), \qquad 10.23$$

where B and ϕ are arbitrary. As we have seen in Sec. 10.1, the term $B\cos(\omega_0 t + \phi)$ is the general solution for the motion of the free undamped oscillator, $m\ddot{x} + kx = 0$. For a damped system, the amplitude B would decrease exponentially in time and eventually we would be left with the steady-state solution

$$x = \frac{F_0}{m}\frac{1}{\omega_0^2 - \omega^2}\cos\omega t.$$

The effects of the initial conditions die out given enough time. In the remainder of this chapter we shall concentrate on the steady-state solution.

Resonance



The amplitude of oscillation, Eq. (10.21), is shown in the sketch as a function of the driving frequency ω. A approaches zero as $\omega \to \infty$ and has a finite value at $\omega = 0$, but it increases without limit at $\omega = \omega_0$, when the oscillator is driven at its natural frequency. This great increase of the amplitude when a system is driven at a certain frequency is known as *resonance*. ω_0 is often called the resonance, or natural, frequency of this system. Equation (10.21) predicts that $A \to \infty$ as $\omega \to \omega_0$, but since no physical system can have infinite amplitude, it is apparent that our solution is inadequate at resonance. The difficulty is due to our neglect of friction; when we take friction into account, we shall see that although the amplitude may be large at resonance, it remains finite.

Equation (10.21) asserts that A is positive for $\omega < \omega_0$ and negative for $\omega > \omega_0$. Negative amplitude means that if the force varies as $\cos\omega t$, the displacement varies as $-\cos\omega t$. Since $-\cos\omega t = \cos(\omega t + \pi)$, the negative sign is equivalent to a phase shift of π radians (i.e., 180°) between the driving force and the

[1] This solution can be verified by direct substitution. In the language of differential equations, the first term on the right in Eq. (10.23) is a particular solution and the second term, $B\cos(\omega t + \phi)$, is the general solution of the homogeneous equation $m\ddot{x} + kx = 0$. These two terms represent the complete solution.

displacement. For $\omega < \omega_0$, the displacement is in phase with the driving force. This phase change through resonance of 180°, which is characteristic of all oscillating systems, is easily demonstrated.

Example 10.4 Forced Harmonic Oscillator Demonstration

Break a long rubber band and suspend something like a heavy pocket knife from one end, holding the other end in your hand. The resonant frequency ω_0 is easily determined by observing the free motion. Now slowly jiggle your hand at a frequency $\omega < \omega_0$: the weight will move in phase with your hand. If you jiggle the system with $\omega > \omega_0$, you will find that the weight always moves in the opposite direction to your hand. For a given amplitude of motion of your hand, the weight moves with decreasing amplitude as ω is increased above ω_0. If you try to jiggle the system at resonance $\omega = \omega_0$, the amplitude increases so much that the weight either flies up in the air or hits your hand. In either case the system no longer behaves like a simple oscillator.

The phenomenon of resonance has both positive and negative aspects in practice. By operating at the resonance frequency of a system we can obtain a response of large amplitude for a very small driving force. Organ pipes utilize this principle effectively, and resonant electric circuits enable us to tune our radios to the desired frequency. On the negative side, we do not want motions of large amplitude in the springs of an automobile or in the crankshaft of its engine. To reduce response at resonance a dissipative friction force is needed. We turn now to the analysis of the forced damped oscillator.

The Forced Damped Harmonic Oscillator

If the motion of the oscillating mass is opposed by a viscous retarding force $-bv$, the total force is

$$\begin{aligned} F &= F_{\text{spring}} + F_{\text{viscous}} + F_{\text{driving}} \\ &= -kx - bv + F_0 \cos \omega t \end{aligned}$$

and the equation of motion can be written

$$m\ddot{x} + b\dot{x} + kx = F_0 \cos \omega t.$$

Dividing by m and using $\gamma = b/m$, $\omega_0^2 = k/m$, we have the standard form

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = \frac{F_0}{m} \cos \omega t. \qquad 10.24$$

To find the steady-state solution we could again try the trick of taking $x = A \cos \omega t$. However, the term $\gamma \dot{x}$ introduces a term in $\sin \omega t$ which does not appear on the right hand side, so that this trial solution is not adequate. This suggests that we try $x = B \cos \omega t + C \sin \omega t = A \cos (\omega t + \phi)$. If this is substituted into Eq. (10.23), you will find that the solution indeed fits provided that A and ϕ have the values

$$A = \frac{F_0}{m} \frac{1}{[(\omega_0{}^2 - \omega^2)^2 + (\omega\gamma)^2]^{\frac{1}{2}}},$$

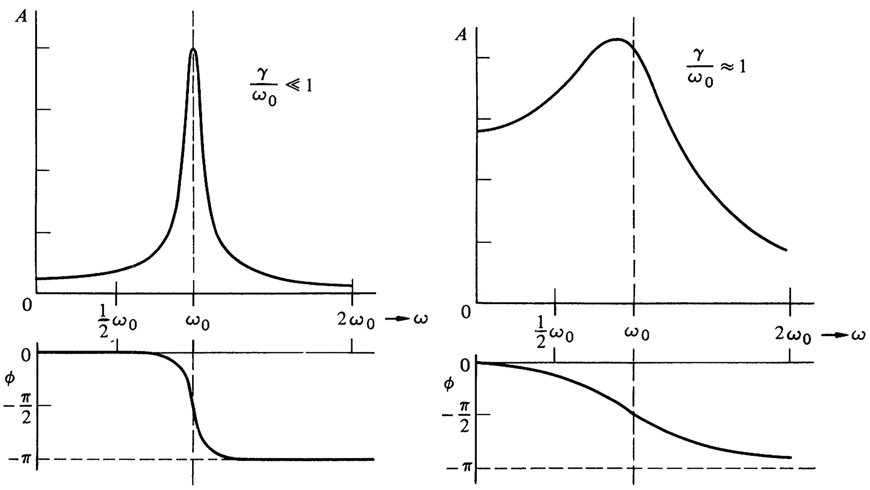
$$\phi = \arctan \left(\frac{\gamma\omega}{\omega^2 - \omega_0{}^2} \right). \qquad 10.25$$

A somewhat more formal method for obtaining this solution is presented in Note 10.2.

The behavior of A and ϕ as functions of ω depends markedly on the ratio γ/ω_0 as the sketches show. For light damping, A is maximum for $\omega = \omega_0$, and the amplitude at resonance is

$$A(\omega_0) = \frac{F_0}{m\omega_0\gamma}.$$

As $\gamma \rightarrow 0$, $A(\omega_0) \rightarrow \infty$, as we expect for an undamped oscillator. Note also that as $\gamma \rightarrow 0$, the phase change occurs more and more abruptly. In the limit $\gamma = 0$, the phase changes from 0 to $-\pi$ when $\omega = \omega_0$.



Resonance in a Lightly Damped System: The Quality Factor Q

Energy considerations simplified our discussion of the undriven damped oscillator in Sec. 10.2, and, similarly, they will be useful to us in the driven case. For the steady-state motion, the amplitude is constant in time. Using

$$x = A\cos(\omega t + \phi) \qquad \text{and} \qquad v = -\omega A \sin(\omega t + \phi),$$

we have

$$K(t) = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m\omega^2A^2\sin^2(\omega t + \phi)$$
$$U(t) = \tfrac{1}{2}kx^2 = \tfrac{1}{2}kA^2\cos^2(\omega t + \phi)$$

and

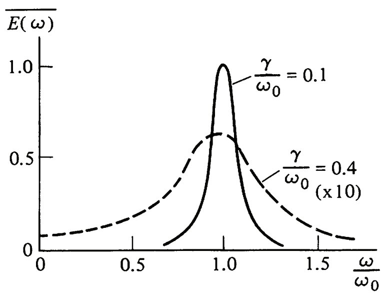
$$\begin{aligned} E(t) &= K(t) + U(t) \\ &= \tfrac{1}{2}A^2[m\omega^2\sin^2(\omega t + \phi) + k\cos^2(\omega t + \phi)]. \end{aligned}$$

The energy is time-dependent and our analysis is simplified if we focus on time average values, as we did in Sec. 10.1. Since $\langle\cos^2(\omega t + \phi)\rangle = \langle\sin^2(\omega t + \phi)\rangle = \frac{1}{2}$, for an average over one period, we have

$$\begin{aligned} \langle K\rangle &= \tfrac{1}{4}m\omega^2A^2 \\ \langle U\rangle &= \tfrac{1}{4}kA^2 \\ \langle E\rangle &= \tfrac{1}{4}A^2(m\omega^2 + k) \\ &= \tfrac{1}{4}mA^2(\omega^2 + \omega_0^2). \end{aligned} \qquad 10.26$$

Let us consider how $\langle E\rangle$ varies as a function of ω. Using Eq. (10.25) for A,

$$\langle E(\omega)\rangle = \frac{1}{4}\frac{F_0^2}{m}\frac{(\omega^2 + \omega_0^2)}{[(\omega_0^2 - \omega^2)^2 + (\omega\gamma)^2]}. \qquad 10.27$$



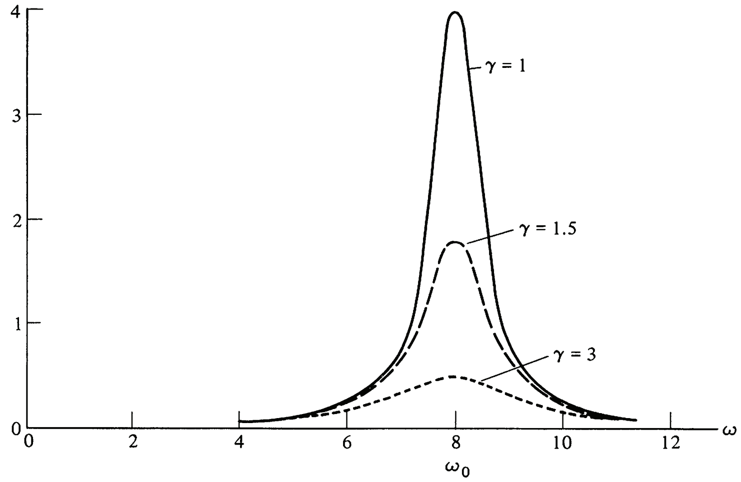
This expression is exact but awkward. It can be written in a much simpler approximate form for the case of light damping, where $\gamma \ll \omega_0$. To see this, consider the sketch of $\langle E(\omega)\rangle$ for $\gamma/\omega_0 = 0.1$ and $\gamma/\omega_0 = 0.4$. For γ sufficiently small, $\langle E(\omega)\rangle$ is effectively zero except near resonance. Hence, there is not much error introduced by replacing ω by ω_0 everywhere in Eq. (10.27) except in the term $(\omega_0^2 - \omega^2)^2$ in the denominator, since this term varies rapidly near resonance. Even this term can be simplified as

$$(\omega_0^2 - \omega^2) = [(\omega_0 + \omega)(\omega_0 - \omega)] \approx 2\omega_0(\omega_0 - \omega).$$

With this approximation, $\langle E(\omega)\rangle$ takes the simple form

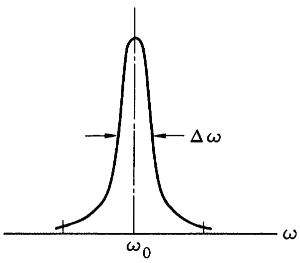
$$\langle E(\omega)\rangle = \frac{1}{4}\frac{F_0^2}{m}\frac{2\omega_0^2}{4\omega_0^2(\omega-\omega_0)^2+\omega_0^2\gamma^2}$$
$$= \frac{1}{8}\frac{F_0^2}{m}\frac{1}{(\omega-\omega_0)^2+(\gamma/2)^2}. \qquad 10.28$$

The plot of the function $[(\omega-\omega_0)^2+(\gamma/2)^2]^{-1}$, which contains the entire frequency dependence of $\langle E(\omega)\rangle$, is called a *resonance curve*, or *lorentzian*. Resonance curves for several values of γ are plotted below. For concreteness, we have taken $\omega_0 = 8$ rad/s. γ is given in units of δ^{-1}.



Let us look more closely at the resonance curve. Its maximum height is $4/\gamma^2$. It falls to one-half maximum when

$$(\omega-\omega_0)^2 = (\gamma/2)^2$$



or when $\omega-\omega_0 = \pm\gamma/2$. The full width of the curve at half maximum value is often called the *resonance width*. If the resonance curve drops to half its maximum value at ω_+ on the high frequency side, and at ω_- on the low frequency side, then the resonance width is $\omega_+ - \omega_- = 2(\gamma/2) = \gamma$. The resonance width is denoted by $\Delta\omega$ in the sketch at left. In general, we have

$$\Delta\omega = \gamma. \qquad 10.29$$

As γ decreases the curve becomes higher and narrower, the range of frequency over which the system responds becomes smaller, and the oscillator becomes increasingly selective in frequency.

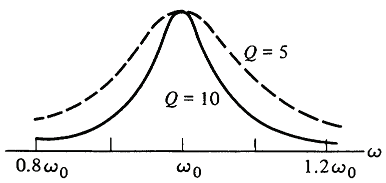
The frequency-selective property of an oscillator can be characterized in a simple fashion by Q, the quality factor introduced in Sec. 10.2. Recall that Q is defined as the ratio of energy stored in the oscillator to energy lost per radian of oscillation. For a lightly damped system oscillating freely, Q has the value

$$Q = \frac{\omega_0}{\gamma},$$

as we showed in Eq. (10.19). The same oscillator, when driven, has a resonance curve with frequency width $\Delta\omega = \gamma$. Hence, the ratio of resonance frequency to the width of the resonance curve, $\omega_0/\Delta\omega$, is $\omega_0/\gamma = Q$. In fact, Q is often defined by

$$Q = \frac{\text{resonance frequency}}{\text{frequency width of resonance curve}}. \qquad 10.30$$

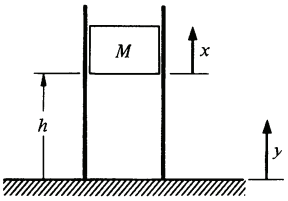
Incidentally, if we had applied the definition of Q in terms of energy to the driven oscillator, the result would have been the same, $Q = \omega_0/\gamma$. The proof of this is left for a problem.



Although Q is fundamentally defined in terms of energy, its chief use in practice is to characterize the frequency response of a system. The drawing shows two resonance curves with different Q's. The heights at resonance have been made equal to facilitate comparison of the widths. It is apparent that the system with $Q = 10$ is considerably more selective than that with $Q = 5$. As pointed out in Example 10.3, certain atomic systems can have a Q greater than 10^8. The sharpness of the resonance curve means that the system will not respond unless driven very near its resonance frequency. Since the resonance frequency is determined by atomic constants, the frequency of oscillation is essentially independent of external influences. Frequencies from such "atomic clocks" are so accurate that they have superseded astronomical time standards.

Example 10.5 Vibration Eliminator

Occasionally one needs to reduce the effect of floor vibrations on a delicate apparatus such as a sensitive balance or a precision optical system. This can be accomplished by mounting the apparatus on an "air



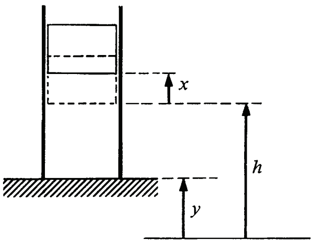
table" whose legs are hollow tubes with pistons supported by air pressure. One such leg is shown schematically in the drawing. The area of the column is A, and the mass it supports is M.

The static forces on M are related by the equilibrium condition

$$P_0A = Mg + P_{at}A,$$

where P_0 is the pressure of gas in the cylinder at equilibrium and P_{at} is the atmospheric pressure on the upper face of M. For some air tables, the weight Mg is much greater than the atmospheric force, and we shall neglect the term $P_{at}A$ in the following. Hence,

$$P_0A = Mg.$$



The equilibrium height of M is h. Let x be the displacement of M from equilibrium relative to an inertial frame. The smaller the value of x, the more nearly motionless the table top will be in inertial space. Floor vibrations cause the lower end of the table leg to move vertically a distance y. When M moves relative to the floor, the volume and the pressure of the trapped gas change. If P is the instantaneous pressure in the cylinder, the equation of motion of M is

$$M\ddot{x} = PA - Mg.$$

According to Boyle's law, the pressure in the cylinder varies inversely with volume for a gas at constant temperature. Therefore

$$\begin{aligned} PV &= \text{constant} \\ &= P_0V_0 \\ &= P_0Ah. \end{aligned}$$

The volume V is

$$V = A(h + x - y).$$

Therefore

$$\begin{aligned} P &= \frac{P_0V_0}{V} = P_0\frac{h}{h + x - y} \\ &\approx P_0\left(1 - \frac{x}{h} + \frac{y}{h}\right)\cdot \end{aligned}$$

In the last step we have assumed that the displacements x and y are small compared with h, the height of the table leg.

The equation of motion becomes

$$M\ddot{x} = P_0A\left(1 - \frac{x}{h} + \frac{y}{h}\right) - Mg.$$

Since we are neglecting the atmospheric force, $P_0A = Mg$, and the equation of motion is simply

$$M\ddot{x} = \frac{Mg}{h}(-x + y)$$

$$\ddot{x} + \frac{g}{h}x = \frac{g}{h}y.$$

If the floor vibration is $y = y_0 \cos \omega t$, M moves like an undamped driven oscillator. Using Eq. (10.22) we see that the solution of the equation is

$$x = x_0 \cos \omega t,$$

where

$$x_0 = y_0 \frac{\omega_0^2}{\omega_0^2 - \omega^2}$$

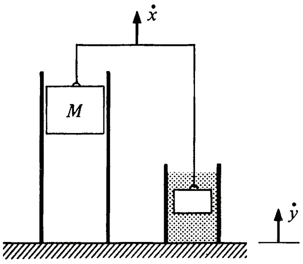
and

$$\omega_0 = \sqrt{\frac{g}{h}}.$$

The object of the air suspension is to make the ratio

$$\frac{x_0}{y_0} = \frac{\omega_0^2}{(\omega_0^2 - \omega^2)}$$

as small as possible. For $\omega \ll \omega_0$, $x_0 = y_0$ and the vibration is transmitted without reduction. For $\omega \gg \omega_0$, $x_0/y_0 = -\omega_0^2/\omega^2$, and the amplitude of vibration is reduced. Thus, for the vibration eliminator to be successful, the resonance frequency must be low compared with the driving frequency. Since $\omega_0 = \sqrt{g/h}$, this requires as long a leg as possible. (Note that the resonance frequency is independent of the mass, a surprising aspect of this type of support.)



The system suffers from one fatal flaw; if vibration occurs near the resonant frequency, the vibration eliminator becomes a vibration amplifier. To avoid this, some damping mechanism must be provided. Often this is accomplished with a device called a *dashpot*, which consists of a piston in a cylinder of oil. The dashpot provides a viscous retarding force $-bv$, where v is the relative velocity of its ends.

$$v = \dot{x} - \dot{y}.$$

The equation of motion is

$$M\ddot{x} = \frac{Mg}{h}(-x + y) - b(\dot{x} - \dot{y})$$

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = \omega_0^2 y + \gamma\dot{y},$$

where

$$\gamma = \frac{b}{M} \qquad \text{and} \qquad \omega_0^2 = \frac{g}{h}.$$

With $y = y_0 \cos \omega t$, this is the equation of a driven damped oscillator. However, the motion of the floor has introduced an additional driving term $\gamma \dot{y} = -\gamma \omega y_0 \sin \omega t$. The steady-state amplitude x_0 can be found by substituting $x = x_0 \cos(\omega t + \phi)$ in the equation. A simpler method is to use complex variables, as outlined in Notes 10.1 and 10.2. Let

$$y = y_0 e^{i\omega t}$$
$$x = x_0 e^{i\omega t}.$$

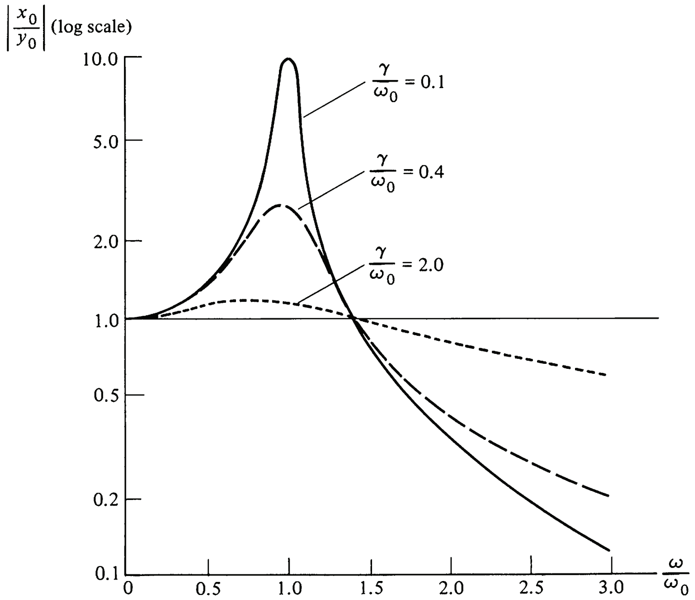
y_0 and x_0 are now complex numbers. Substituting in the equation of motion gives

$$(-\omega^2 + i\omega\gamma + \omega_0^2)x_0 e^{i\omega t} = (\omega_0^2 + i\omega\gamma)y_0 e^{i\omega t}$$

$$x_0 = \left[\frac{\omega_0^2 + i\omega\gamma}{(\omega_0^2 - \omega^2) + i\omega\gamma}\right] y_0.$$

We are interested in the ratio of the magnitudes, $|x_0|/|y_0|$.

$$\frac{|x_0|}{|y_0|} = \sqrt{\frac{x_0 x_0^*}{y_0 y_0^*}}$$
$$= \left[\frac{\omega_0^4 + (\omega\gamma)^2}{(\omega_0^2 - \omega^2)^2 + (\omega\gamma)^2}\right]^{\frac{1}{2}}.$$



The graph shows $|x_0|/|y_0|$ versus ω/ω_0 for various values of γ/ω_0. For ω/ω_0 less than about 1.5, $|x_0|/|y_0| > 1$. The vibration is actually enhanced, showing that even with damping it is essential to reduce the resonance frequency below the driving frequency. When ω/ω_0 is greater than 1.5, $|x_0|/|y_0| < 1$. For these higher frequencies, the vibration isolation is more effective the smaller the damping. However, small damping increases the danger from vibrations near resonance. Practical air tables have resonance frequencies of 1 Hz or less.

Many vibration elimination systems use springs instead of an air suspension. However, this does not change the form of the equation of motion. Often coil springs are used in automobiles to isolate the chassis from road vibrations. Damping is provided by shock absorbers, a type of dashpot. The resonance frequency is $\omega_0 = \sqrt{k/M}$, where k is the spring constant and M is the mass. If a smooth turnpike ride is the chief consideration, one wants a massive car with weak damping and soft springs. Such a car is difficult to control on a bumpy road where resonance can be excited. The best suspensions are heavily damped and feel rather stiff. The danger in driving with defective shock absorbers is that the car may be thrown out of control if it is excited at resonance by bumps in the road.

10.4 Response in Time Versus Response in Frequency

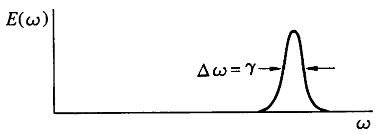
The smaller the damping of a free oscillator, the more slowly its energy is dissipated. The same oscillator, when driven, becomes increasingly more frequency selective as the damping is decreased. As we shall now show, the time dependence of the free oscillator and the frequency dependence of the driven oscillator are intimately related.

Recall from Eq. (10.16) that the energy of a free oscillator is

$$E(t) = E_0 e^{-\gamma t}.$$

The damping time is $\tau = 1/\gamma$.

Next, consider the response in frequency of the same oscillator when it is driven by a force $F_0 \cos \omega t$. From Eq. (10.29) the width of the resonance curve is



Forced oscillator

$$\Delta\omega = \gamma.$$

The damping time τ and the resonance curve width $\Delta\omega$ obey

$$\tau\,\Delta\omega = 1. \qquad 10.31$$

According to this result it is impossible to design an oscillator with arbitrary damping time and resonance width; if we choose one, the other is automatically fixed by Eq. (10.31).

Equation (10.31) has many implications for the design of mechanical and electrical systems. Any element which is highly frequency selective will oscillate for a long time if it is accidentally perturbed. Furthermore, such an element will take a long time to reach the steady state when a driving force is applied because the effects of the initial conditions die out only slowly. More generally, Eq. (10.31) plays a fundamental role in quantum mechanics; it is closely related to one form of the Heisenberg uncertainty principle.

Note 10.1 Solution of the Equation of Motion for the Undriven Damped Oscillator

THE USE OF COMPLEX VARIABLES

All the equations of motion in this chapter can be solved simply by using complex variables.[1] Here is a summary of the algebra of complex numbers.

1. Every complex number z can be written in the cartesian form $x + iy$, where $i^2 = -1$. x is the *real* part of z, and y is the *imaginary* part. The sum of two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is the complex number $z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$. The product of z_1 and z_2 is

$$\begin{aligned} z_1z_2 &= (x_1 + iy_1)(x_2 + iy_2) \\ &= x_1x_2 + ix_1y_2 + iy_1x_2 + i^2y_1y_2 \\ &= (x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2). \end{aligned}$$

If two complex numbers are equal, the real parts and the imaginary parts are respectively equal.

$$x_1 + iy_1 = x_2 + iy_2$$

implies that

$$\begin{aligned} x_1 &= x_2 \\ y_1 &= y_2. \end{aligned}$$

2. $z^* \equiv x - iy$ is the *complex conjugate* of $z = x + iy$. The quantity $|z| = \sqrt{zz^*}$ is the *magnitude* of z.

$$\begin{aligned} |z| &= \sqrt{zz^*} \\ &= [(x + iy)(x - iy)]^{\frac{1}{2}} \\ &= \sqrt{x^2 + y^2}. \end{aligned}$$

[1] A simple treatment of the algebra of complex numbers may be found in most of the calculus texts listed at the end of Chap. 1.

3. Every complex number z can be written in the polar form $re^{i\theta}$. r is a real number, the *modulus*, and θ is the *argument*. To go from cartesian to polar form we use De Moivre's theorem

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Hence,

$$\begin{aligned} re^{i\theta} &= r\cos\theta + ir\sin\theta \\ &= x + iy, \end{aligned}$$

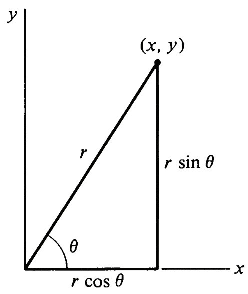
from which it follows that

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned}$$

and

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \arctan\frac{y}{x}. \end{aligned}$$

We see that $r = |z|$.

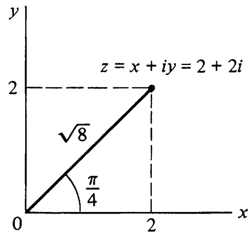


Complex numbers can be represented graphically. Let the horizontal axis be the real (x) axis, and the vertical axis be the imaginary (y) axis. The complex number $x + iy$ is represented by the point (x,y). As the sketch shows, introduction of the polar form is analogous to the use of plane polar coordinates.

Here are some examples:

1. Express $z = (3 + 4i)/(2 + i)$ in cartesian form. The method is to multiply numerator and denominator by the complex conjugate of the denominator.

$$\begin{aligned} z &= \frac{3+4i}{2+i} \\ &= \frac{3+4i}{2+i}\cdot\frac{2-i}{2-i} \\ &= \frac{6+8i-3i-4i^2}{4+2i-2i-i^2} \\ &= \frac{10+5i}{5} \\ &= 2+i \end{aligned}$$



2. Express $z = 2 + 2i$ in polar form.

$$r = |z| = \sqrt{2^2 + 2^2} = \sqrt{8}$$

$$\theta = \arctan\frac{2}{2} = \frac{\pi}{4}$$

$$z = \sqrt{8}e^{i\pi/4}$$

THE DAMPED OSCILLATOR

We turn now to the equation for the damped oscillator.

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0 \qquad 1$$

To cast this into complex form we introduce the companion equation

$$\ddot{y} + \gamma\dot{y} + \omega_0^2 y = 0. \qquad 2$$

Multiplying Eq. (2) by i and adding it to Eq. (1) yields

$$\ddot{z} + \gamma\dot{z} + \omega_0^2 z = 0. \qquad 3$$

Note that either the real or imaginary part of z is an acceptable solution for the equation of motion.

Since the coefficients of the derivatives of z are all constants, a natural choice for the solution of Eq. (3) is

$$z = z_0 e^{\alpha t}, \qquad 4$$

where z_0 and α are independent of t. With this trial solution Eq. (3) yields

$$\alpha^2 z_0 e^{\alpha t} + \alpha\gamma z_0 e^{\alpha t} + \omega_0^2 z_0 e^{\alpha t} = 0.$$

Dividing out the common factor $z_0 e^{\alpha t}$, we have

$$\alpha^2 + \alpha\gamma + \omega_0^2 = 0, \qquad 5$$

which has the solution

$$\alpha = -\frac{\gamma}{2} \pm \sqrt{\frac{\gamma^2}{4} - \omega_0^2}. \qquad 6$$

Let us call the two roots α_1 and α_2. We see that our solution can be written as

$$z = z_A e^{\alpha_1 t} + z_B e^{\alpha_2 t},$$

where z_A and z_B are constants.

There are three possible forms of the solution, depending on whether α is real or complex. We consider these solutions in turn.

Case 1 Light Damping: $\gamma^2 \ll 4\omega_0^2$

In this case $\sqrt{\gamma^2/4 - \omega_0^2}$ is imaginary and we can write

$$\alpha = -\frac{\gamma}{2} \pm i\sqrt{\omega_0^2 - \frac{\gamma^2}{4}} \qquad 7$$

$$= -\frac{\gamma}{2} \pm i\omega_1,$$

where

$$\omega_1 = \sqrt{\omega_0^2 - \frac{\gamma^2}{4}}.$$

The solution is

$$z = e^{-(\gamma/2)t}(z_1 e^{i\omega_1 t} + z_2 e^{-i\omega_1 t}),$$

where z_1 and z_2 are complex constants. In order to find the real part of z we write the complex numbers in cartesian form.

$$x + iy = e^{-(\gamma/2)t}[(x_1 + iy_1)(\cos\omega_1 t + i\sin\omega_1 t) + (x_2 + iy_2)(\cos\omega_1 t - i\sin\omega_1 t)]$$

The real part x is

$$x = e^{-(\gamma/2)t}(B\cos\omega_1 t + C\sin\omega_1 t)$$

or

$$x = Ae^{-(\gamma/2)t}\cos(\omega_1 t + \phi), \qquad 8$$

where A and ϕ are new arbitrary constants. This is the result quoted in Eq. (10.9). Incidentally, the imaginary part of z, which is also an acceptable solution, has exactly the same form.

Case 2 Heavy Damping: $\gamma^2/4 > \omega_0^2$

In this case, $\sqrt{\gamma^2/4 - \omega_0^2}$ is real and Eq. (5) has the solution

$$\alpha = -\frac{\gamma}{2} \pm \frac{\gamma}{2}\sqrt{1 - \frac{4\omega_0^2}{\gamma^2}}.$$

Both roots are negative, and we have

$$z = z_1 e^{-|\alpha_1|t} + z_2 e^{-|\alpha_2|t}. \qquad 9$$

The exponentials are real. The real part of z is

$$x = Ae^{-|\alpha_1|t} + Be^{-|\alpha_2|t}. \qquad 10$$

This solution has no oscillatory behavior; the motion is known as *overdamped.*

Case 3 Critical Damping: $\gamma^2/4 = \omega_0^2$

If $\gamma^2/4 = \omega_0^2$ we have only the single root

$$\alpha = -\frac{\gamma}{2}.$$

The corresponding solution is

$$x = Ae^{-(\gamma/2)t}. \qquad 11$$

However, this solution is incomplete. Mathematically, the solution of a second order linear differential equation always involves two arbitrary constants. Physically, the solution must have two constants to allow us to specify the initial position and initial velocity of the oscillator. As described in texts on differential equations, the second solution can be found by using a "variation of parameters" trial solution.

$$x = u(t)e^{(-\gamma/2)(t)}.$$

Substituting in Eq. (1) and recalling that $\gamma = 2\omega_0$ for this case, we find that $u(t)$ must satisfy the equation

$$\ddot{u} = 0.$$

Hence,

$$u = a + bt$$

and the general solution is

$$x = Ae^{-(\gamma/2)t} + Bte^{-(\gamma/2)t}. \qquad 12$$

Note 10.2 Solution of the Equation of Motion for the Forced Oscillator

We wish to solve

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = \frac{F_0}{m}\cos\omega t. \qquad 1$$

Consider the companion equation

$$\ddot{y} + \gamma\dot{y} + \omega_0^2 y = \frac{F_0}{m}\sin\omega t. \qquad 2$$

Multiplying Eq. (2) by i and adding to Eq. (1) yields

$$\ddot{z} + \gamma\dot{z} + \omega_0^2 z = \frac{F_0}{m}e^{i\omega t}. \qquad 3$$

z must vary as $e^{i\omega t}$, so we try

$$z = z_0 e^{i\omega t}.$$

Inserting this in Eq. (3) gives

$$(-\omega^2 + i\omega\gamma + \omega_0^2)z_0 e^{i\omega t} = \frac{F_0}{m} e^{i\omega t}$$

or

$$z_0 = \frac{F_0}{m} \frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma}.$$

We can put z_0 into cartesian form by multiplying numerator and denominator by the complex conjugate of the denominator.

$$z_0 = \frac{F_0}{m} \frac{1}{(\omega_0^2 - \omega^2) + i\omega\gamma} \frac{(\omega_0^2 - \omega^2) - i\omega\gamma}{(\omega_0^2 - \omega^2) - i\omega\gamma}$$
$$= \frac{F_0}{m} \frac{(\omega_0^2 - \omega^2) - i\omega\gamma}{(\omega_0^2 - \omega^2)^2 + (\omega\gamma)^2}$$

In polar form, $z_0 = Re^{i\phi}$, where

$$R = \sqrt{z_0 z_0^*}$$
$$= \frac{F_0}{m} \left[\frac{1}{(\omega_0^2 - \omega^2)^2 + (\omega\gamma)^2}\right]^{\frac{1}{2}} \qquad 4$$

and

$$\phi = \arctan\left(\frac{\omega\gamma}{\omega^2 - \omega_0^2}\right). \qquad 5$$

The complete solution is

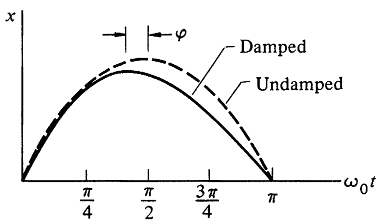
$$z = Re^{i\phi}e^{i\omega t},$$

which has the real part

$$x = R\cos(\omega t + \phi).$$

The steady-state motion is completely specified by the amplitude R and the phase angle ϕ. Both R and ϕ are contained implicitly in the single complex number z_0.

Problems



10.1 Show by direct calculation that $\langle \sin^2(\omega t)\rangle = \frac{1}{2}$, where the time average is taken over any complete period $t_1 \le t \le t_1 + 2\pi/\omega$.

Show also that $\langle \sin(\omega t)\cos(\omega t)\rangle = 0$ when the average is over a complete period.

10.2 A 0.3-kg mass is attached to a spring and oscillates at 2 Hz with a Q of 60. Find the spring constant and damping constant.

10.3 In an undamped free harmonic oscillator the motion is given by $x = A\sin\omega_0 t$. The displacement is maximum exactly midway between the zero crossings.

In a damped oscillator the motion is no longer sinusoidal, and the maximum is advanced before the midpoint of the zero crossings. Show that the maximum is advanced by a phase angle ϕ given approximately by

$$\phi = \frac{1}{2Q},$$

where we assume that Q is large.

10.4 The *logarithmic decrement* δ is defined to be the natural logarithm of the ratio of successive maximum displacements (in the same direction) of a free damped oscillator. Show that $\delta = \pi/Q$.

Find the spring constant k and damping constant b of a damped oscillator having a mass of 5 kg, frequency of oscillation 0.5 Hz, and logarithmic decrement 0.02.

10.5 If the damping constant of a free oscillator is given by $\gamma = 2\omega_0$, the system is said to be critically damped. Show by direct substitution that in this case the motion is given by

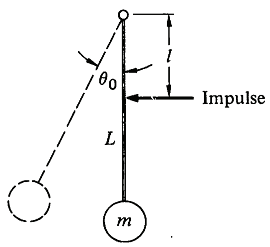
$$x = (A + Bt)e^{-(\gamma/2)t},$$

where A and B are constants.

A critically damped oscillator is at rest at equilibrium. At $t = 0$ it is given a blow of total impulse I. Sketch the motion, and find the time at which the displacement starts to decrease.

10.6 *a.* A mass of 10 kg falls 50 cm onto the platform of a spring scale, and sticks. The platform eventually comes to rest 10 cm below its initial position. The mass of the platform is 2 kg. Find the spring constant.

b. It is desired to put in a damping system so that the scale comes to rest in minimum time without overshoot. This means that the scale must be critically damped (see Note 10.1). Find the necessary damping constant and the equation for the motion of the platform after the mass hits.



10.7 Find the driving frequency for which the velocity of a forced damped oscillator is exactly in phase with the driving force.

10.8 The pendulum of a grandfather's clock activates an escapement mechanism every time it passes through the vertical. The escapement is under tension (provided by a hanging weight) and gives the pendulum a small impulse a distance l from the pivot. The energy transferred by this impulse compensates for the energy dissipated by friction, so that the pendulum swings with a constant amplitude.

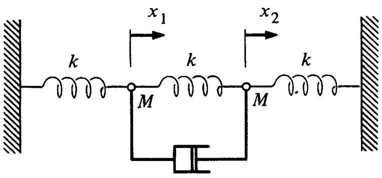
a. What is the impulse needed to sustain the motion of a pendulum of length L and mass m, with an amplitude of swing θ_0 and quality factor Q?

b. Why is it desirable for the pendulum to engage the escapement as it passes vertical rather than at some other point of the cycle?

10.9 Show that for a lightly damped forced oscillator near resonance

$$\frac{\text{average energy stored in the oscillator}}{\text{average energy dissipated per radian}} \approx \frac{\omega_0}{\gamma} = Q.$$

10.10 A small cuckoo clock has a pendulum 25 cm long with a mass of 10 g and a period of 1 s. The clock is powered by a 200-g weight which falls 2 m between the daily windings. The amplitude of the swing is 0.2 rad. What is the Q of the clock? How long would the clock run if it were powered by a battery with 1 J capacity?



10.11 Two particles, each of mass M, are hung between three identical springs. Each spring is massless and has spring constant k. Neglect gravity. The masses are connected as shown to a dashpot of negligible mass.

The dashpot exerts a force bv, where v is the relative velocity of its two ends. The force opposes the motion. Let x_1 and x_2 be the displacements of the two masses from equilibrium.

a. Find the equation of motion for each mass.

b. Show that the equations of motion can be solved in terms of the new dependent variables $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$.

c. Show that if the masses are initially at rest and mass 1 is given an initial velocity v_0, the motion of the masses after a sufficiently long time is

$$x_1 = x_2$$
$$= \frac{v_0}{2\omega} \sin \omega t.$$

Evaluate ω.

10.12 The motion of a damped oscillator driven by an applied force $F_0 \cos \omega t$ is given by $x_a(t) = A \cos (\omega t + \phi)$, where A and ϕ are given by Eq. (10.25). Consider an oscillator which is released from rest at $t = 0$. Its motion must satisfy $x(0) = 0$, $v(0) = 0$, but after a very long time, we expect that $x(t) = x_a(t)$. To satisfy these conditions we can take as the solution

$$x(t) = x_a(t) + x_b(t),$$

where $x_b(t)$ is the solution to the equation motion of the free damped oscillator, Eq. (10.8).

a. Show that if $x_a(t)$ satisfies the equation of motion for the forced damped oscillator, then so does $x(t) = x_a(t) + x_b(t)$, where $x_b(t)$ satisfies the equation of motion of the free damped oscillator, Eq. (10.25).

b. Choose the arbitrary constants in $x_b(t)$ so that $x(t)$ satisfies the initial conditions. [$x_b(t)$ is given by Eq. (10.9). Note that A and ϕ here are arbitrary.]

c. Sketch the resulting motion for the case where the oscillator is driven at resonance.

11 THE SPECIAL THEORY OF RELATIVITY

11.1 The Need for a New Mode of Thought

In some ways the structure of physics resembles a mansion whose outward form is apparent to the casual visitor but whose inner life —the customs and rituals which give a special outlook and kinship to its occupants—require time and effort to comprehend. Indeed, initiation into this special knowledge is the goal of our present endeavor. In the first ten chapters we introduced and applied the fundamental laws of classical mechanics; hopefully you now feel familiar with these laws and have come to appreciate their beauty, their essential simplicity, and their power.

Unfortunately, in order to present dynamics in a concise and tidy form, we have generally sidestepped discussion of how physics actually grew. In Chaps. 11 through 14 we are going to discuss one of the great achievements of modern physics, the special theory of relativity. Rather than present the theory as a completed structure—a simple set of postulates with the rules for their application—we shall depart from our previous style and look into the background of the theory and its rationale.

If the structure of physics is a mansion, it is a mansion of ancient origin. It is founded on the remains of prehistoric hovels where man first kept track of the moon and tried to understand the simple patterns of nature. Traces of antiquity lie hidden in the site: Phoenician and Egyptian, Babylonian, and, of course, Greek. Compass and straightedge lie scattered among lodestone and amber, artifacts of astrologer and alchemist. The mansion is built on the debris of false starts and painful struggles to understand nature honestly. None of this is visible, however, and we take the present structure much for granted. The outer shell was built in the seventeenth century by Kepler, Galileo, Newton, and others, such as Huygens, Hooke, Leibniz, Bernoulli, and Boyle. The major architects have one characteristic in common: while extending the external dimensions of the mansion by applying physics to new areas, they also deepened its foundations by advancing our knowledge of the fundamental laws. The greatest of these figures is Newton, who revealed the laws of dynamics and of gravity, cornerstones of modern science. At the same time he vigorously applied physics to the natural world. Newton executed meticulous experiments in heat flow, optics, and the motion of bodies under viscous forces; he investigated the shape of the moon, the tides along the coast of England, and how to build bridges.

The momentum generated by Newton's discoveries gave physics an impetus which is still very much with us. The eighteenth and nineteenth centuries saw a flowering of science as physicists such as Euler, Lagrange, Laplace, Faraday, and Maxwell extended our knowledge of the physical world. However, their efforts were directed at upward extension of the mansion; Newton's account of the fundamental laws of physics was so overwhelming, and so successful, that not until the last quarter of the nineteenth century was there a serious attempt to investigate the foundations.

It was the German physicist Ernst Mach who first successfully challenged newtonian thought. Although Mach's work left newtonian physics more or less intact, his thinking was crucial in the revolution shortly to come. In 1883 Mach published his text "The Science of Mechanics," which incorporated a critique of newtonian physics, the first incisive criticism of Newton's theory of dynamics. In addition to presenting a lucid account of newtonian mechanics, the text incorporates several significant contributions to the fundamentals of mechanics. Mach clarified newtonian dynamics by carefully analyzing Newton's explanation of the dynamical laws, taking care to distinguish between definitions, derived results, and statements of physical law. Mach's approach is now widely accepted; our discussion of Newton's laws in Chap. 2 is very much in Mach's spirit.

"The Science of Mechanics" raised the question of the distinction between absolute and relative motion. Mach pointed out Newton's ambivalence on this subject, although he went on to show that the question was irrelevant to the application of newtonian dynamics. In the process he dwelt on the problem of inertia and enunciated the principle that now bears his name: inertia is not an intrinsic property of matter or space but depends on the existence of all matter in the universe. We encountered Mach's principle in our discussion of fictitious forces in Chap. 8, but we shall not dwell on it here for it turns out that the problem of inertia was not the crucial difficulty with newtonian mechanics.

The fundamental weakness in newtonian dynamics, as Mach pointed out, centers on Newton's conception of space and time. In a preface to his dynamical theory, Newton avowed that he would forgo abstract speculation and deal only with observable facts. Although such a point of view is now commonplace, at the time it represented a tremendous intellectual leap. Before Newton, the business of natural philosophy was to explain the reasons for things, to find a rational account for the workings of nature, rather than to describe natural phenomena quantitatively. Newton essentially reversed the priorities. Against the criticism that his theory of universal gravitation merely described gravity without accounting for its origin, Newton replied "I do not make hypotheses."

Unfortunately, Newton was not completely faithful to his resolve to avoid abstract speculation and to deal only with demonstrable facts. In particular, consider the following description of time that appears in the "Principia." (The excerpt is condensed.)

> Absolute, true and mathematical time, of itself and by its own true nature, flows uniformly on, without regard to anything external.
>
> Relative, apparent and common time is some sensible and external measure of absolute time estimated by the motions of bodies, whether accurate or inequable, and is commonly employed in place of true time; as an hour, a day, a month, a year.

Mach comments that "it would appear as though Newton in the remarks cited here still stood under the influence of medieval philosophy, as though he had grown unfaithful to his resolve to investigate only actual facts." Mach goes on to point out that since time is necessarily measured by the repetitive motion of some physical system, for instance the pendulum of a clock or the revolution of the earth about the sun, then the properties of time must be connected with the laws which describe the motions of physical systems. Simply put, Newton's idea of time without clocks is metaphysical; to understand the properties of time we must observe the properties of clocks. A simple idea? Yes, indeed, were it not for the fact that the idea of absolute time is so natural that the eventual consequences of Mach's position, the relativistic description of time, still come as something of a shock to the student of science.

There are similar weaknesses in the newtonian view of space. Mach argued that since position in space is determined with measuring rods, the properties of space can be understood only by investigating the properties of meter sticks. We must look to nature to understand space, not to platonic ideals.

Mach's special contribution was to examine the most elemental aspects of newtonian thought, to look critically at matters which seem too simple to discuss, and to insist that we turn to experience to understand the properties of nature rather than to rely on abstractions of the mind. From this point of view, Newton's assumptions about space and time must be regarded merely as postulates. Classical mechanics follows from these postulates, but other assumptions are possible and from them different laws of dynamics could follow.

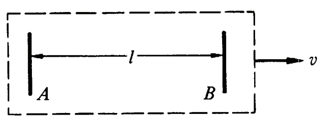
Mach's critique had little immediate effect, but its influence was eventually profound. In particular, the youthful Einstein, while a student at the Polytechnic Institute in Zurich in the period 1897–1900, was much attracted by Mach's ideas on the foundations of newtonian physics and by Mach's insistence that physical concepts be defined in terms of observables. However, the immediate cause for the overthrow of newtonian physics was not Mach's criticisms of newtonian thought. The difficulties lay with Maxwell's electromagnetic theory, the crowning achievement of classical physics. Traditionally, the problem is presented in terms of a single crucial experiment that decisively condemned classical physics, the Michelson-Morley experiment, and most treatments of special relativity take this experiment as the point of departure. We shall follow this tradition, but we should point out that history is not that simple. In the first place, Albert A. Michelson, who conceived and executed the experiment, never regarded it as crucial. Michelson viewed the experiment as a flop for not giving the expected result, a view he maintained long after its full significance became known. Furthermore, it now appears that the Michelson-Morley experiment played little, if any, role in Einstein's thinking. In fact, there is good reason to believe that Einstein knew nothing of the experiment until after he had published his theory of relativity in 1906. Nevertheless, the Michelson-Morley experiment so clearly dramatizes the essential dilemma of electromagnetic theory that we shall bow to tradition and take it as our starting point.

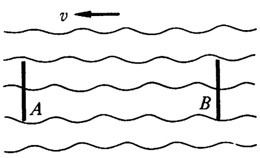
11.2 The Michelson-Morley Experiment

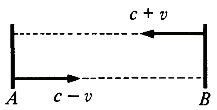
The problem to which Michelson devoted himself was that of determining the effect of the earth's motion on the velocity of light. Briefly, Maxwell's electromagnetic theory (1861) predicted that electromagnetic disturbances in empty space would propagate at 3×10^8 m/s—the speed of light. The simplest disturbance is a periodic wave, and the evidence was overwhelming that light consisted of electromagnetic waves. However, there were conceptual difficulties.

The only waves previously known to physics were mechanical waves propagating in solids, liquids, and gases. A sound wave in air, for example, consists of alternate regions of higher and lower pressure propagating with a speed of 330 m/s, somewhat less than the speed of molecular motion. The speed of mechanical waves in metals is higher, typically 5,000 m/s, and increases with the strength of the "spring forces" between neighboring atoms.

Electromagnetic wave propagation seemed to be a very different sort of thing. The ether, the medium which supposedly supported the electromagnetic disturbance, had to be immensely rigid to give a speed of 3×10^8 m/s. At the same time it had to be insubstantial enough not to interfere with the motion of the planets. Maxwell's theory itself made no essential reference to the ether, but Maxwell and his contemporaries were unable to accept the idea of waves propagating in empty space.







The speed of a sound wave v_s depends on the properties of the medium. If we observe a sound wave from a coordinate system moving relative to the medium, the speed of sound will appear to be greater or less than v_s, depending on whether we move in the direction of propagation or against it. Similarly, Maxwell pointed out that the speed of the earth as it circled the sun, 3×10^4 m/s, should change the apparent speed of light.

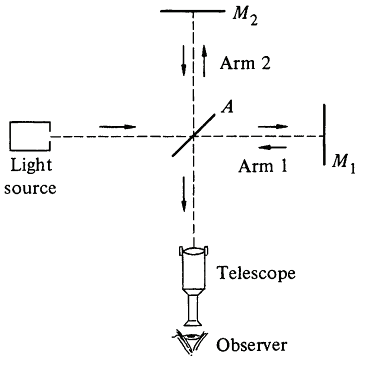
Suppose that light makes a round trip ABA between two points A and B separated by distance l. The apparatus is moving through the ether to the right, as shown in the upper drawing. Relative to the apparatus, the ether is moving to the left, as shown in the second drawing. The velocity of light relative to the apparatus is $c + v$ to the left, and $c - v$ to the right.

The transit time from A to B is $t_1 = l/(c - v)$, and from B to A it is $t_2 = l/(c + v)$. If the apparatus were at rest, t_1 and t_2 would have the value $t_0 = l/c$. The effect of the earth's motion is to delay the return of the light signal by

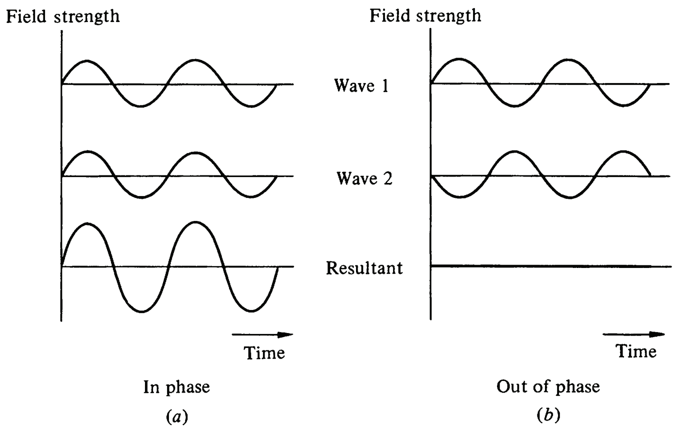
$$\begin{aligned}
\Delta t &= t_1 + t_2 - 2t_0 \\
&= \frac{l}{c - v} + \frac{l}{c + v} - 2\frac{l}{c} \\
&= \frac{l}{c}\left(\frac{1}{1 - v/c} + \frac{1}{1 + v/c} - 2\right) \\
&= 2\frac{l}{c}\left(\frac{1}{1 - v^2/c^2} - 1\right) \\
&\approx 2\frac{l}{c}\frac{v^2}{c^2}.
\end{aligned}$$

For the earth in orbit $v/c = 10^{-4}$, and if we take l to be typical of a laboratory apparatus, $l = 1$ m, then $\Delta t = 2 \times 1/(3 \times 10^8) \times$

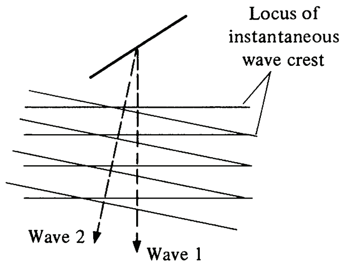
$10^{-8} \approx 7 \times 10^{-17}$ s, an interval much too small to be measured directly. Fortunately, Michelson was not discouraged. In 1881 he came up with the following solution.

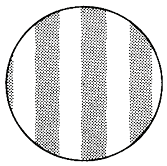


Rather than measure the time of transit of one light beam, Michelson observed the *difference* between the transit times of two beams. His device is sketched at the left. The light from the source is split into two beams by a thinly silvered mirror, A. Half the light passes through mirror A to mirror M_1, where it is reflected back to mirror A and then to the observer. The other half of the light from the source is diverted up the second arm and strikes mirror M_2, which reflects it to the observer. If the two arms are identical, the light waves recombine at mirror A just as if they had never separated: the observer sees an illuminated field of view. The situation is drastically altered if either beam suffers a delay. Suppose, for instance, that beam 1 is delayed by exactly one-half cycle of oscillation. The waves arrive in opposite phase and exactly cancel each other: the observer's field is dark.

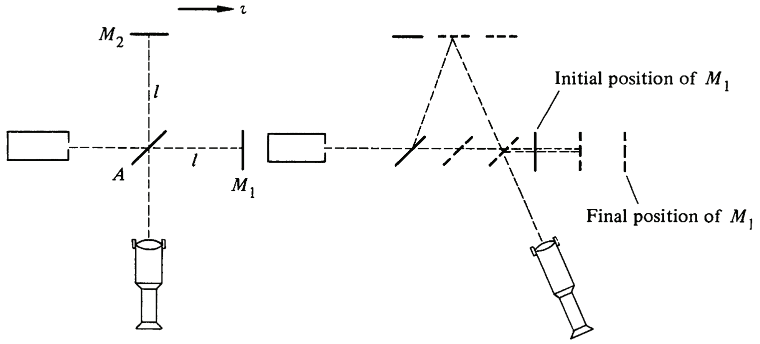


The two cases are shown in the sketches above. The vertical displacement corresponds to the strength of the electric field of light at the observer's eye. The fields of the two beams add vectorially. For visible light the period of the wave is typically 10^{-15} s, too fast for our eyes to follow. Rather, our eyes respond to the average power of the wave which is proportional to the square of the resultant field. Thus, beams in phase, sketch (*a*), give steady bright illumination, and beams out of phase, sketch (*b*), give darkness.





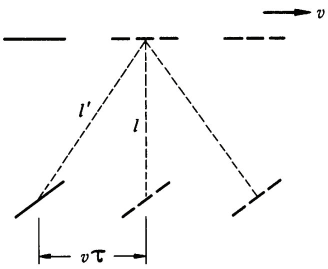
Usually one of the mirrors is slightly tilted. This produces a gradual time delay across the returning wavefront, as shown in the first sketch, and the two interfering waves go in and out of phase across the field of view. The observer sees alternate light and dark bands, as in the second sketch. If the length of either arm is changed, the fringe pattern shifts; a change in path of one wavelength shifts the pattern by one fringe. Since the light traverses each arm twice, once in each direction, a change in the length of either arm by one-half wavelength produces a shift of one fringe. With care it is possible to measure a small fraction of a fringe shift; one can readily observe a path change of one-hundredth wavelength, approximately 10^{-8} m. (Michelson also used his interferometer to measure the length of the standard meter bar; he essentially created the field of high precision measurement.)



Suppose that the interferometer is oriented so that one axis lies along the direction of motion of the earth, as shown. The time for the wave to travel from mirror A to mirror M_1 and back is

$$\begin{aligned} T_1 &= \frac{l}{c-v} + \frac{l}{c+v} \\ &= \frac{l}{c}\left(\frac{1}{1-v/c} + \frac{1}{1+v/c}\right) \\ &= \frac{2l}{c}\left(\frac{1}{1-v^2/c^2}\right) \approx \frac{2l}{c}\left(1+\frac{v^2}{c^2}\right), \end{aligned}$$

where l is the length of the arm. There is also a time delay along arm 2, but this is a trifle more subtle to calculate. (Michelson overlooked it in the first report of his experiment in 1882.) For



the beam to return to its initial point on the thinly silvered mirror, it must traverse the angular path shown at left. Let τ be the time it takes the wavefront to go from mirror A to mirror M_2. The distance actually traversed is $l' = (l^2 + v^2\tau^2)^{\frac{1}{2}}$ and, since $l' = c\tau$ we have

$$\tau = \frac{(l^2 + v^2\tau^2)^{\frac{1}{2}}}{c}$$

or

$$\tau^2 = \frac{l^2}{c^2} + \frac{v^2}{c^2}\tau^2.$$

It follows that

$$\tau = \frac{l}{c}\frac{1}{\sqrt{1 - v^2/c^2}}.$$

The time for the wave to travel from mirror A to mirror M_2 and back is

$$\begin{aligned} T_2 &= 2\tau \\ &= 2\frac{l}{c}\frac{1}{\sqrt{1 - v^2/c^2}} \\ &\approx 2\frac{l}{c}\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right). \end{aligned}$$

The difference between the travel times of the beams is

$$\begin{aligned} \Delta T &= T_1 - T_2 \\ &= \frac{l}{c}\frac{v^2}{c^2}. \end{aligned}$$

The delay ΔT shifts the fringe pattern from where it would be if the earth were at rest. However, there is a major problem: the fringe scale has no "zero," since the arms cannot be made identical in length to the needed accuracy. Michelson hit upon the idea of watching the fringes as the apparatus is rotated by 90°. The rotation effectively interchanges arms 1 and 2. The change in the delay between the two positions is $2\Delta T$, and the corresponding fringe shift is readily calculated. If λ is the wavelength of the illuminating light, a time delay of λ/c will shift the pattern by one fringe. Thus, the time delay $2\Delta T$ will shift the pattern N fringes, where

$$N = \frac{2\Delta T}{(\lambda/c)} = \frac{2l}{\lambda}\frac{v^2}{c^2}.$$

If the arms have unequal lengths, l_1 and l_2, this result still holds, provided that we replace $2l$ by $l_1 + l_2$.

In Michelson's first apparatus, the arm length was 1.2 m, or, as he put it, 2×10^6 wavelengths of yellow (sodium) light. Since $v/c = 10^{-4}$, we expect

$$N = 2(2 \times 10^6)(10^{-4})^2 = 0.04.$$

Although this is not a large shift, Michelson had adequate resolution to see it. To his disappointment, he found no measurable shift in the fringe pattern. A much more refined experiment, executed with E. W. Morley, in 1887, used multiple reflections to increase the expected shift to 0.4 fringe. Although a shift as small as 0.01 fringe could have been detected, no effect was seen. The experiment has been repeated many times since, but always with negative results. It appears that we are unable to detect our motion through the ether.

11.3 The Postulates of Special Relativity

The elusive nature of the ether presented physics with a troublesome enigma. Maxwell attempted to devise a mechanical model of the ether, but as he continued to develop his theory of light, the ether played a less and less important role, until finally it was altogether absent. The ether vanished like the Cheshire Cat, leaving only a smile behind. After the Michelson-Morley experiment, even the smile had vanished. Numerous attempts to explain the null results of the Michelson-Morley experiment introduced such complexity as to threaten the foundations of electromagnetic theory. The most successful attempt was the hypothesis suggested independently by FitzGerald and by Lorentz that motion of the earth through the ether caused a shortening of one arm of the Michelson interferometer by exactly the amount required to eliminate the fringe shift. However, their speculations were based on an assumed model of atomic forces, and even though they arrived at some of the formulas shortly to be obtained by Einstein, their reasoning was far less general. Other theories which involved such artifacts as drag of the ether by the earth were even less productive.

The Universal Velocity

It is an indication of Einstein's genius that the troublesome problem of the ether pointed the way not to complexity and elaboration but to a simplification that unified the basic concepts of physics. Einstein viewed the difficulty with the ether not as stemming from a fault of electromagnetic theory but as arising from an error in basic dynamical principles. He argued that since the velocity of light predicted by electromagnetic theory, c, involves no reference to a medium, c must be a universal constant, the same for all observers. Thus, if we measure the speed of light from a source, the result will always be c, independent of our motion. This is in marked contrast to the case of sound waves, for example, where the observed speed depends on motion of the observer with respect to the medium. The ideas of a universal velocity was indeed a bold hypothesis, contrary to all previous experience and, for many of Einstein's contemporaries, defying common sense. But common sense is often a poor guide. Einstein once remarked that common sense consists of all the prejudices one learns before the age of eighteen.

The Principle of Relativity

The special theory of relativity involves one additional postulate—the assertion that the laws of physics have the same form with respect to all inertial systems. This principle, known as the principle of relativity, was not novel; Galileo is credited with first pointing out that there is no way to determine whether one is moving uniformly or is at rest, and Newton, although troubled by this point, gave it a rigorous expression in his dynamical laws in which acceleration, not velocity, is paramount. The principle of relativity played only a minor role in the development of classical mechanics; Einstein elevated it to a keystone of dynamics. He extended the principle to include not only the laws of mechanics but also the laws of electromagnetic interaction and, by supposition, all the laws of physics. Furthermore, in his hands the principle of relativity became an important working principle in discovering the correct form of physical laws. We can only surmise the sources of his inspiration, but they must have included the following consideration. If the velocity of light were not a universal constant, that is, if the ether could be detected, then the principle of relativity would fail; a special inertial frame would be singled out, the one at rest in the ether. However, the form of Maxwell's equations, as well as the failure of any experiment to detect motion through the ether, suggests that the speed of light is constant, independent of the motion of the source. Our inability to detect absolute motion, either with light or with newtonian forces, implies that absolute motion has no role in physics.

Whereas most physicists regarded the absence of the ether as a paradox, Einstein saw that its absence preserved the simplicity of the principle of relativity. His view was essentially conservative; he insisted on preserving the principle of relativity which the ether would destroy. Apparently the urge toward simplicity was fundamental to his personality.[1] The special theory of relativity was the simplest way to preserve the unity of classical physics. In fact, as we shall see in the closing chapter, special relativity actually simplifies newtonian thought by combining space and time in a natural fashion from which the various conservation laws follow as a single entity.

The Postulates of Special Relativity

To summarize, the postulates of special relativity are:

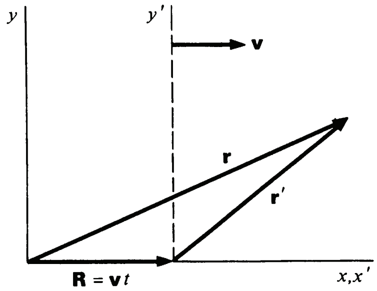
The laws of physics have the same form in all inertial systems.

The velocity of light in empty space is a universal constant, the same for all observers.

The mathematical expression of the special theory of relativity is embodied in the Lorentz transformations—a simple prescription for relating events in different inertial systems. Contrary to the mystique, the mathematics of relativity is quite simple: elementary algebra will suffice. The reasoning is also simple, but it has a deceptive simplicity. We start by looking once more at the Galilean transformations.

[1] Einstein had much in common with Newton. In the second book of his "Principia," Newton states his rules of scientific reasoning. Rule 1 is: "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. . . . Nature is pleased with simplicity. . ."

11.4 The Galilean Transformations



Let us review for a moment the newtonian way of viewing an event in different coordinate systems. Consider an inertial system x, y, z, in which we are at rest, and a second inertial system x', y', z', which is translating uniformly in the $+x$ direction with velocity v. For convenience, we take the origins to coincide at $t = 0$, and take the axes to be parallel.

If a particular point in space has coordinates $\mathbf{r} = (x,y,z)$ in our "rest" system, the corresponding coordinates in the moving system are $\mathbf{r}' = (x',y',z')$. These are related by

$$\mathbf{r}' = \mathbf{r} - \mathbf{R},$$

where

$$\mathbf{R} = \mathbf{v}t.$$

Since v is in the x direction, we have

$$\begin{aligned} x' &= x - vt \\ y' &= y \\ z' &= z \\ t' &= t. \end{aligned} \qquad 11.1$$

The last equation is listed merely for completeness. It follows from the newtonian idea of an "absolute" time, and it is so taken for granted that it is generally omitted in discussions of classical physics.

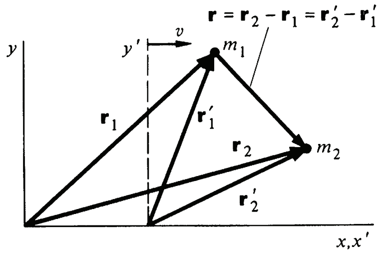
Equations (11.1) are known as the *Galilean transformations.* Since the laws of newtonian mechanics hold in all inertial systems, they are unaffected by these transformations. The classical principle of relativity asserts that the laws of mechanics are unchanged by the Galilean transformations. The following example illustrates the meaning of this statement.

Example 11.1 **The Galilean Transformations**

Consider how we might discover the law of force between two isolated bodies from observations of their motion. For example, the problem might be to discover the law of gravitation from data on the elliptical orbit of one of Jupiter's moons. If m_1 and m_2 are the masses of the moon and of Jupiter, respectively, and $\mathbf{r}_1$ and $\mathbf{r}_2$ are their positions relative to an astronomer on the earth, we have

$$m_1\ddot{\mathbf{r}}_1 = \mathbf{F}(r)$$

$$m_2\ddot{\mathbf{r}}_2 = -\mathbf{F}(r),$$



where we assume that **F**, the force between the bodies, depends only on their separation $r = |\mathbf{r}_1 - \mathbf{r}_2|$. (Including the effect of the sun makes the analysis more cumbersome without changing the conclusions.)

From our data on $\mathbf{r}_1(t)$ we can evaluate $\ddot{\mathbf{r}}_1$, which yields the value of **F**, (or $\mathbf{F}/m_1$, to be more precise). In principle, this is the procedure Newton followed in discovering the law of universal gravitation. Suppose that the data show $\mathbf{F}(r) = -Gm_1m_2\hat{\mathbf{r}}/r^2$.

Now let us consider the problem from the point of view of an astronomer in a spacecraft observatory which is flying by the earth. According to the principle of relativity he must obtain the same force law. The situation is represented in the drawing. x, y is the earthbound system, x', y' is the spacecraft system, and v is the relative velocity.

In the x', y' system the astronomer concludes that the force on m_1 is

$$\mathbf{F}'(r') = m_1\ddot{\mathbf{r}}_1'.$$

However,

$$\mathbf{r}_1 = \mathbf{r}_1' + \mathbf{v}t$$
$$\dot{\mathbf{r}}_1 = \dot{\mathbf{r}}_1' + \mathbf{v}$$
$$\ddot{\mathbf{r}}_1 = \ddot{\mathbf{r}}_1'.$$

Hence,

$$\begin{aligned}\mathbf{F}'(r') &= m_1\ddot{\mathbf{r}}_1' \\ &= m_1\ddot{\mathbf{r}}_1 \\ &= \mathbf{F}(r).\end{aligned}$$

Since $r' = r$, $\mathbf{F}'(r') = \mathbf{F}'(r)$. But we have just shown that $\mathbf{F}'(r') = \mathbf{F}(r)$. Hence,

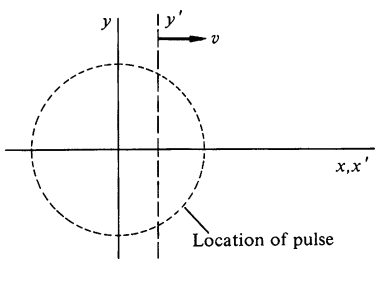
$$\begin{aligned}\mathbf{F}'(r) &= \mathbf{F}(r) \\ &= -\frac{Gm_1m_2}{r^2}\hat{\mathbf{r}}.\end{aligned}$$

The law of force is identical to the one found on earth. This is what we mean when we say that the two inertial systems are equivalent. If the form of the law, or the value of G, were different in the two systems, we could make a judgment about the speed of a coordinate system by investigating the law of gravitation in that system. The systems would not be equivalent.

Example 11.1 is almost trivial, since the force depends on the separation of the two particles, a quantity which is unchanged (invariant) under the Galilean transformations. In newtonian physics, all forces are due to interactions between particles, interactions which depend on the *relative* coordinates of the particles. Consequently they are invariant under the Galilean transformations.

What happens to the equation for a light signal under the Galilean transformations? The following example shows the difficulty that arises.

Example 11.2 **A Light Pulse as Described by the Galilean Transformations**



At $t = 0$ a pulse of light is emitted isotropically in the x, y system. It travels outward with velocity c. The equation for the wavefront along the x axis is

$$x = ct.$$

In the x', y' system, the equation for the wavefront along the x' axis is

$$\begin{aligned} x' &= x - vt \\ &= (c - v)t, \end{aligned}$$

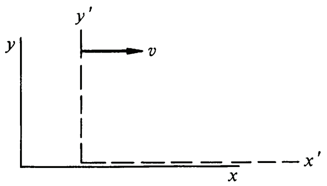
where v is the relative velocity of the two systems.

The x' velocity of the pulse in the x', y' system is

$$\frac{dx'}{dt} = c - v.$$

But this is contrary to the postulate that the speed of light is a universal constant c for all observers. Clearly, the Galilean transformations are inadequate.

11.5 The Lorentz Transformations



Since the Galilean transformations do not satisfy the postulate that the speed of light is a universal constant, Einstein proposed an alternate prescription for describing the same event in different inertial systems. Let us refer once more to our standard systems, the rest system, x, y, z, t and the system x', y', z', t' which moves with velocity v along the positive x axis. The origins coincide at $t = t' = 0$. We take the most general transformation relating the coordinates of a given event in the two systems to be of the form

$$x' = Ax + Bt \qquad 11.2a$$

$$y' = y \qquad 11.2b$$

$$z' = z \qquad 11.2c$$

$$t' = Cx + Dt. \qquad 11.2d$$

The transformations are linear, for otherwise there would not be a simple one-to-one relation between events in the different systems. For instance, a nonlinear transformation would predict acceleration in one system even if the velocity were constant in

TABLE 11.1

EVENT	COOR-DINATES (x,y,t)	COOR-DINATES (x',y',t')	TRANSFORMATION LAW	RESULT
Observer in (x,y) sees origin of (x',y') move along x axis with velocity v.	$x = vt$	$x' = 0$	$x' = Ax + Bt$ 11.2a $0 = Avt + Bt$	$B = -Av$
Observer in (x',y') sees origin of (x,y) move along x' axis with velocity $-v$.	$x = 0$	$x' = -vt'$	$x' = A(x - vt)$ 11.2a $t' = Cx + Dt$ 11.2d $A(0 - vt) = -v(0 + Dt)$	$D = A$
A light pulse is sent out from origin along x axis at $t = 0$. Its location is given by:	$x = ct$	$x' = ct'$	$x' = A(x - vt)$ 11.2a $t' = Cx + At$ 11.2d $A(ct - vt) = c(Cct + At)$	$C = -\frac{Av}{c^2}$
A light pulse is emitted along the y axis in (x,y) at $t = 0$. In (x',y') the pulse has components along the x' and y' axes. The velocity of the pulse is c in both systems. Its coordinates are given by:	$x = 0$ $y = ct$	$x'^2 + y'^2 = c^2t'^2$	$x' = A(x - vt)$ 11.2a $y' = y$ 11.2b $t' = A(-vx/c^2 + t)$ 11.2d $A^2(0 - vt)^2 + (ct)^2 = c^2A^2[-(v/c^2)0 + t]^2$	†$A = \frac{1}{\sqrt{1 - v^2/c^2}}$

† In general, $A = \pm 1/\sqrt{1 - v^2/c^2}$. We choose the positive root; otherwise, in the limit $v = 0$ we would find $x' = -x$ rather than $x' = x$ as we require.

the other, clearly an unacceptable property for a transformation between inertial systems. We have assumed that the y' and z' axes are left unchanged by the transformation for reasons of symmetry, which we shall discuss later.

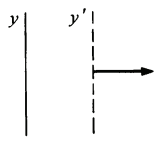
Equations (11.2) contain four unknown constants. To evaluate these we consider four cases in which we know *a priori* how an event appears in the two systems. This is carried out in Table 11.1.

Inserting the results of Table 11.1 into Eq. (11.2) gives

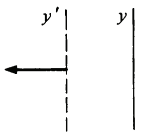
$$\begin{aligned} x' &= \frac{1}{\sqrt{1 - v^2/c^2}}(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \frac{1}{\sqrt{1 - v^2/c^2}}\left(t - \frac{vx}{c^2}\right) \end{aligned} \qquad 11.3$$

It is a straightforward matter to solve these equations algebraically for x, y, z, t in terms of x', y', z', t'. Alternatively, we can simply interchange the labels and reverse the sign of v, because the only difference between the systems is the direction of the relative velocity. The result is

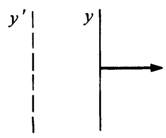
$$\begin{aligned} x &= \frac{1}{\sqrt{1 - v^2/c^2}}(x' + vt') \\ y &= y' \\ z &= z' \\ t &= \frac{1}{\sqrt{1 - v^2/c^2}}\left(t' + \frac{vx'}{c^2}\right) \end{aligned} \qquad 11.4$$



(a)



(b)



(c)

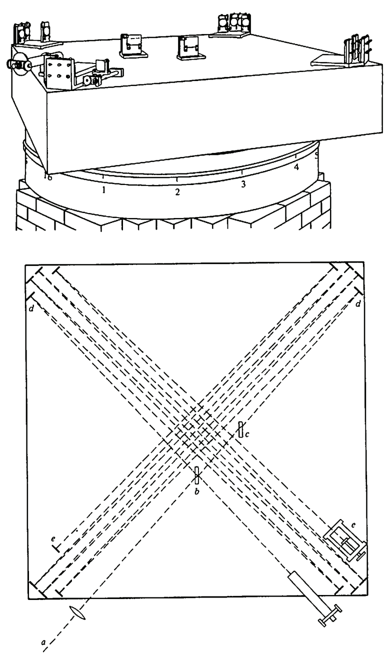
Equations (11.3) and (11.4) are the *Lorentz transformations*, the prescription for relating the coordinates of an event in different inertial systems so as to satisfy the postulates of special relativity. In the following chapters we shall explore their consequences. We conclude the present discussion by explaining the argument for assuming $y = y'$, $z = z'$.

Consider a section of the y and y' axes as shown in figure (a). The y' axis is moving to the right with velocity v.

If we look at the systems from behind the paper, the situation appears as in sketch (b).

Since only relative motion is important, Figure (b) is equivalent to (c). However, (c) is identical to (a) except that y' and y are interchanged. We conclude that the y and y' axes are indistinguishable and $y = y'$. By a similar argument $z = z'$.

Problems



(From *American Journal of Science*, November, 1887.)

11.1 The Michelson-Morley experiment was carried out at the Case School of Applied Science (now Case-Western Reserve University) in 1887. The apparatus was a refined version of the interferometer used by Michelson in his initial search in Berlin during 1881. The interferometer was mounted on a granite slab 5 ft square and 14 in thick resting on a float riding in a mercury-filled trough. The effective length of the interferometer arms was lengthened to 11 m by the use of mirrors. The light source was the yellow line of sodium, $\lambda = 590 \times 10^{-9}$ m. Michelson and Morley found no systematic shift of fringe with direction, although they could have detected a shift as small as one-hundredth fringe.

How does the upper limit to the earth's velocity through the ether set by this experiment compare with the earth's orbital velocity around the sun, 30 km/s? See drawing at left.

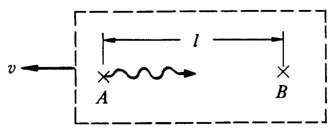
11.2 If the two arms of a Michelson interferometer have lengths l_1 and l_2, show that the fringe shift when the interferometer is rotated by 90° with respect to the velocity v through the ether is

$$N = \frac{l_1 + l_2}{\lambda}\frac{v^2}{c^2},$$

where λ is the wavelength of the light.

11.3 The Irish physicist G. F. FitzGerald and the Dutch physicist H. A. Lorentz independently tried to explain the null result of the Michelson-Morley experiment by the following hypothesis: motion of a body through the ether sets up a strain which causes the body to contract along the line of motion by the factor $1 - \frac{1}{2}v^2/c^2$. Show that this hypothesis accounts for the absence of a fringe shift in the Michelson-Morley experiment. (The hypothesis was disproved in 1932 by R. J. Kennedy and E. M. Thorndike, who repeated the Michelson-Morley experiment with an interferometer having arms of different lengths.)

11.4 The Michelson-Morley experiment is known as a second order experiment because the observed effect depends on $(v/c)^2$. Consider the following first order experiment.



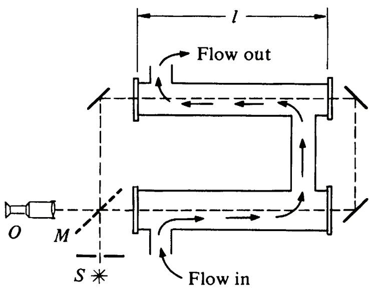
At time $t = 0$, observer A sends a signal to observer B a distance l away. B records the arrival time. Assume that the system is moving through the ether with speed v in the direction shown. Suppose that the laboratory is then rotated 180° with respect to the velocity, reversing the positions of A and B. At time $t = T$, A sends a second signal to B.

a. Show that the interval B observes between the arrival of the signals is $T + \Delta T$, where

$$\Delta T = \frac{2l}{c}\frac{v}{c}$$

to order $(v/c)^3$.

b. Assume that the experiment is done between a clock on the ground and one in a satellite overhead. For an orbit with a 24-h period, $l = 5.6R_e$, where R_e is the earth's radius. Present atomic clocks approach a stability of 1 part in 10^{14}. What is the smallest value of v that this experiment could detect using such clocks?



11.5 In 1851 H. L. Fizeau investigated the velocity of light through a moving medium using the interferometer shown. Light of wavelength λ from a source S is split into two beams by the mirror M. The beams travel around the interferometer in opposite directions and are combined at the telescope of the observer, O, who sees a fringe pattern. Two arms of the interferometer pass through water-filled tubes of length l with flat glass end plates. The water runs through the tubes, so that one of the light beams travels downstream while the other goes upstream. The velocity of light in water at rest is c/n, where n is the refractive index of water. If we assume that the velocity of the water is added to the velocity of light in the downstream direction, and subtracted in the upstream direction, show that the fringe shift which occurs when the water flows with velocity v is

$$N = 4n^2 \frac{l}{\lambda c} v.$$

(The actual fringe shift measured by Fizeau was

$$N = \frac{4n^2 l}{\lambda c} fv,$$

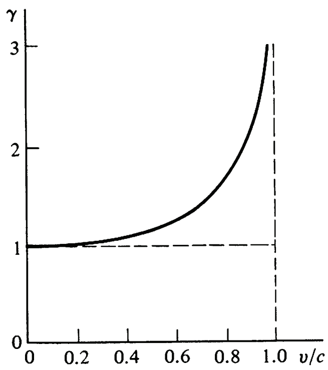
where $f = 1 - 1/n^2$. f, known as the Fresnel drag coefficient, was postulated in 1818, but it was not satisfactorily explained until the advent of relativity. It is derived in the next chapter.)

12 RELATIVISTIC KINEMATICS

12.1 Introduction

The special theory of relativity demands that we examine and modify the familiar results of newtonian physics. We must start by reconsidering kinematics, the most elementary aspect of mechanics, a topic apparently so simple that we gave little thought to its foundations in our earlier discussion. As we pointed out in the last chapter, classical kinematics obeys the Galilean transformations. We must now develop the kinematics appropriate to the Lorentz transformations.

The Lorentz transformations are simplified by introducing



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

Since $(v/c)^2 \leq 1$, γ is greater than or equal to one. The Lorentz transformations, Eqs. (11.3) and (11.4), then take the form

$$\begin{aligned} x' &= \gamma(x - vt) & x &= \gamma(x' + vt') \\ y' &= y & y &= y' \\ z' &= z & z &= z' \\ t' &= \gamma\left(t - \frac{xv}{c^2}\right) & t &= \gamma\left(t' + \frac{x'v}{c^2}\right). \end{aligned} \qquad 12.1$$

It is important to understand clearly the function of the Lorentz transformations, for the lore of relativity is filled with so-called paradoxes (generally simple mistakes) in which the Lorentz transformations are misapplied and lead to contradictory results. The Lorentz transformations relate the coordinates of a *single event* in one inertial system to the coordinates of the *same event* in a second inertial system. Examples of single events are:

A light pulse leaves the point $x = 3$ m, $y = 7$ m, $z = -4$ m at $t = 5$ s.

The origin of the x', y', z' system passes the origin of the x, y, z system at time t.

One end of a stick lies at the point x', y', z', at time t'.

A bearer of evil tidings bursts into the king's chamber at midnight.

Single events are characterized by a set of definite values for the coordinates x, y, z, t. More complicated events can be described by a collection of single events. For example, consider a stick lying along the y axis. The location of the stick is defined by *two* single events—the coordinates of its end points at a particular time.

Before setting out to apply the Lorentz transformations, we should consider carefully how to determine the coordinates of an event. Often we speak of "an observer"; for instance, "an observer in the x', y' system sees a flash of light at $x' = 1$, $y' = 3$, $t' = 0$." This is a handy way to describe observations, but there are conceptual difficulties with the idea of a single observer. Consider an observer who notes that a pulse of light leaves the origin at $t = 0$, and finds that at time t_A the pulse is at $x_A = ct_A$. To make such an observation he would have to move to position x_A before the light arrived there—he would have to move faster than light. As we shall see, this is impossible. However, it is nevertheless possible to record the coordinates of any series of events we please by assuming that we have many observers stationed throughout space. Each one has his own clock, and each is assigned to a specific location, x, y, z. Every time an event occurs at a particular location, the local observer notes the time. Later, all the observers send reports to a central office which prepares a complete record of the times and locations of all events in the system. When we talk of "an observer," we mean someone who has, at least in principle, a copy of this record.

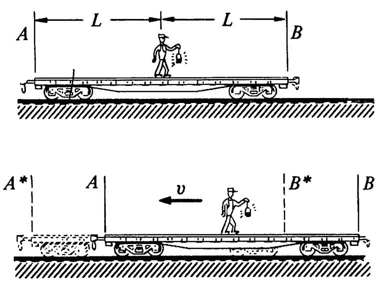
In order for the procedure to work it is essential that all the clocks run at the same rate and that they be synchronized. There is a subtle point here, for synchronized clocks will not appear to agree unless they are at the same location. For example, suppose that we use a powerful telescope to look at a clock on the moon. Since it takes light approximately 1 s to travel from the moon to the earth, a moon clock should indicate 1 s before noon when an earth clock indicates noon, provided that the two clocks are properly synchronized. Similarly, the earth clock should appear to be 1 s late to an observer on the moon. By extension, this procedure can be used to synchronize all the clocks in a particular inertial system.

12.2 Simultaneity and the Order of Events

We have an intuitive idea of what is meant when we say that two events are simultaneous. With respect to a given coordinate system, two events are simultaneous if their time coordinates have the same value. However, as the following example shows, events which are simultaneous in one coordinate system are not necessarily simultaneous in a second coordinate system.

Example 12.1 Simultaneity

Consider a railwayman standing at the middle of a freight car of length $2L$. He flicks on his lantern and a light pulse travels out in all directions



with the velocity c. Light arrives at the two ends of the car after a time interval L/c. In this system, the freight car's rest system, the light arrives simultaneously at A and B.

Now let us observe the same situation from a different frame, one moving to the right with velocity v. In this frame the freight car moves to the left with velocity v. As observed in this frame the light still has velocity c, according to the second postulate of special relativity. However, during the transit time, A moves to A^* and B moves to B^*. It is apparent that the pulse arrives at B^* before A^*; the events are not simultaneous in this frame.

Example 12.1 shows that once we accept the postulates of relativity we are forced to abandon the intuitive idea of simultaneity. The Lorentz transformations, which embody the postulates of relativity, allow us to calculate the times of events in two different systems.

Example 12.2 An Application of the Lorentz Transformations

How do we find the time of arrival of the light pulse at each end of the freight car in the last example? The problem is trivial in the rest frame. Take the origin of coordinates at the center of the car, and take $t = 0$ at the instant the lantern flashes. The two events are

Event 1:

Pulse arrives at end A $\begin{cases} x_1 = -L \\ t_1 = \dfrac{L}{c} = T \end{cases}$

Event 2:

Pulse arrives at end B $\begin{cases} x_2 = L \\ t_2 = \dfrac{L}{c} = T \end{cases}$

To find the time of the events in the moving system we apply the Lorentz transformations for the time coordinates.

Event 1:

$$\begin{aligned} t_1' &= \gamma\left(t_1 - \frac{vx_1}{c^2}\right) \\ &= \gamma\left(T + \frac{vL}{c^2}\right) \\ &= \frac{1}{\sqrt{1 - v^2/c^2}}\left(T + \frac{v}{c}T\right) \\ &= T\sqrt{\frac{1 + v/c}{1 - v/c}}. \end{aligned}$$

Event 2:

$$t_2' = \gamma\left(t_2 - \frac{vx_2}{c^2}\right)$$
$$= T\sqrt{\frac{1 - v/c}{1 + v/c}}.$$

In the moving system, the pulse arrives at B (event 2) earlier than it arrives at A, as we anticipated.

As we saw in the last two examples, simultaneity is not a particularly fundamental property of events; it depends on the coordinate system. Is it possible to find a coordinate system in which any two events are simultaneous? As the following example shows, there are two classes of events. For two given events, we can either find a coordinate system in which the events are simultaneous or one in which the events occur at the same point in space.

Example 12.3 The Order of Events: Timelike and Spacelike Intervals

Two events A and B have the following coordinates in the x, y system.

Event A:

x_A, t_A.

Event B:

x_B, t_B.

(For both events, $y = 0$.)

The distance L and time T separating the events in the x, y system are

$$L = x_B - x_A$$
$$T = t_B - t_A.$$

For concreteness, we take L and T to be positive. To find the coordinates in the x', y' system we use the Lorentz transformations, Eq. (12.1):

$$x_A' = \gamma(x_A - vt_A)$$
$$t_A' = \gamma\left(t_A - \frac{vx_A}{c^2}\right)$$
$$x_B' = \gamma(x_B - vt_B)$$
$$t_B' = \gamma\left(t_B - \frac{vx_B}{c^2}\right).$$

The distance L' between the events in the x', y' system is

$$L' = x'_B - x'_A$$
$$= \gamma[x_B - x_A - v(t_B - t_A)]$$
$$L' = \gamma(L - vT).$$

Similarly,

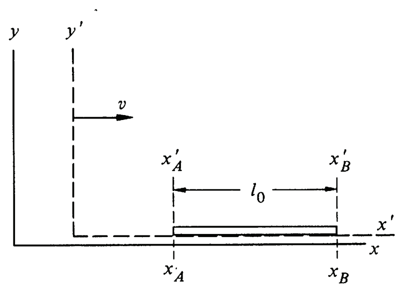
$$T' = \gamma\left(T - \frac{vL}{c^2}\right).$$

Assuming that v is always less than c, it follows that if $L > cT$, L' is always positive, while T' can be positive, negative, or zero. Such an interval is called *spacelike,* since it is possible to choose a system in which the events are simultaneous, namely, a system moving with $v = c^2T/L$. On the other hand, if $L < cT$, T' is always positive, whereas L' can be positive, negative, or zero. The interval is then known as *timelike,* since it is possible to find a coordinate system in which the events occur at the same point.

12.3 The Lorentz Contraction and Time Dilation

Two dramatic results of the special theory of relativity are that a meter stick is shorter when moving than when it is at rest, and that a moving clock runs slow. These results are quite real: the experimental evidence for relativity is so overwhelming that physicists now regard such kinematic effects as commonplace.

The Lorentz Contraction



Consider a stick at rest in the x', y' system, lying along the x' axis with its ends at x'_A and x'_B. The length of the stick is $l_0 = x'_B - x'_A$. l_0 is called the "rest," or "proper," length of the stick: it is what we normally mean when we talk of length. The system x', y' is called the rest, or proper, system of the stick.

Now let us determine the length of the stick l in the system in which the observer is at rest. This system, known as the "laboratory" system, has coordinates x, y. In the laboratory system the stick moves to the right with velocity v.

The length of a stick is the distance between its ends at the same instant of time. The end points must be determined simultaneously in the *lab* system; we must find the correspondence between x' and x at some value of t. This is readily accomplished by applying the Lorentz transformation $x' = \gamma(x - vt)$. We have

$$x'_B = \gamma(x_B - vt)$$
$$x'_A = \gamma(x_A - vt).$$

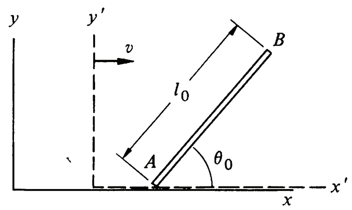
Subtracting, we obtain $l_0 = \gamma l$, or

$$l = \frac{l_0}{\gamma} = l_0\sqrt{1 - \frac{v^2}{c^2}}.$$

l is shorter than l_0: the meter stick is contracted. As $v \to c$, $l \to 0$. This shortening, known as the Lorentz contraction, occurs only along the direction of motion: if the stick lay along the y axis, we would use the transformation $y' = y$ to find $l_0 = l$.

A word of caution. The following argument is fallacious—but it is easy to get trapped by it. "In the rest system, the end of the stick has coordinates x'_A and x'_B at some time $t' = 0$. To find the length in the lab system we use $x = \gamma(x' + vt')$, and obtain $l = \gamma l_0$. Hence, the moving stick looks long." The error is that the end points must be measured simultaneously in the *lab* system. These measurements will not be simultaneous in the rest system, but this is of no consequence.

Example 12.4 The Orientation of a Moving Rod



A rod of length l_0 lies in the $x'y'$ plane of its rest system and makes an angle θ_0 with the x' axis. What is the length and orientation of the rod in the lab system x, y in which the rod moves to the right with velocity v?

Designate the ends of the rod A and B. In the rest system these points have coordinates

$$A: \quad x'_A = 0 \qquad y'_A = 0$$
$$B: \quad x'_B = l_0 \cos\theta_0 \qquad y'_B = l_0 \sin\theta_0.$$

We require the coordinates of A and B in the lab system at a time t. We use $x' = \gamma(x - vt)$, $y' = y$ to obtain:

$$A: \quad x'_A = 0 = \gamma(x_A - vt) \qquad y'_A = 0 = y_A$$
$$B: \quad x'_B = l_0 \cos\theta_0 = \gamma(x_B - vt) \qquad y'_B = l_0 \sin\theta_0 = y_B.$$

Hence,

$$x_B - x_A = \frac{l_0 \cos\theta_0}{\gamma}$$
$$y_B - y_A = l_0 \sin\theta_0.$$

The length is

$$\begin{aligned} l &= [(x_B - x_A)^2 + (y_B - y_A)^2]^{\frac{1}{2}} \\ &= l_0\left[\left(1 - \frac{v^2}{c^2}\right)\cos^2\theta_0 + \sin^2\theta_0\right]^{\frac{1}{2}} \\ &= l_0\left[1 - \frac{v^2}{c^2}\cos^2\theta_0\right]^{\frac{1}{2}}. \end{aligned}$$

The angle that the rod makes with the x axis is

$$\begin{aligned}\theta &= \arctan \frac{y_B - y_A}{x_B - x_A} \\ &= \arctan\left(\gamma \frac{\sin \theta_0}{\cos \theta_0}\right) \\ &= \arctan (\gamma \tan \theta_0).\end{aligned}$$

The moving rod is both contracted and rotated.

Time Dilation

Next we investigate the effect of motion on time. Consider a clock at rest in the x', y' system and consider two events A and B, both occurring at the same point x_0':

$$\begin{aligned}&A: \quad x_0' \quad t_A' \\ &B: \quad x_0' \quad t_B'.\end{aligned}$$

The interval $\tau = t_B' - t_A'$ is the time interval between the events in the rest system. It is called the *proper time* interval.

In order to find the corresponding time interval in the laboratory system we use $t = \gamma(t' + x'v/c^2)$.

$$\begin{aligned}t_A &= \gamma\left(t_A' + \frac{vx_0'}{c^2}\right) \\ t_B &= \gamma\left(t_B' + \frac{vx_0'}{c^2}\right).\end{aligned}$$

Subtracting to obtain $T = t_B - t_A$, we find

$$\begin{aligned}T &= \gamma(t_B' - t_A') \\ &= \gamma\tau \\ &= \frac{\tau}{\sqrt{1 - v^2/c^2}}.\end{aligned}$$

The time interval in the laboratory system is greater than that in the rest system; the moving clock runs slow. This effect, known as *time dilation*, has important practical consequences.

Example 12.5 Time Dilation and Meson Decay

The lifetime of cosmic ray μ mesons (muons) has become a classic demonstration of time dilation. The effect was first observed by B. Rossi and

D. B. Hall[1] and is the subject of an excellent film by D. H. Frisch and J. H. Smith.[2]

The experiment hinges on the fact that the muon is an unstable particle which spontaneously decays into an electron and two neutrinos. The meson carries either a positive or negative charge and decays into either a positive electron (positron, e^+) or ordinary electron (e^-).

Symbolically, we can write

$$\mu^\pm \rightarrow e^\pm + \nu + \bar{\nu}.$$

ν stands for neutrino and $\bar{\nu}$ for antineutrino. The decay of the μ meson is typical of radioactive decay processes: if there are $N(0)$ muons at $t = 0$, the number at time t is

$$N(t) = N(0)e^{-t/\tau},$$

where τ, the mean lifetime, is 2.15×10^{-6} s. Muons can be observed by stopping them in dense absorbers and detecting the decay electron, which comes off with an energy of about 40 MeV (1 MeV = 1 million electronvolts = 1.6×10^{-13} J).

μ mesons are formed in abundance when high energy cosmic ray protons enter the earth's upper atmosphere. The protons lose energy rapidly, and at the altitude of a typical mountaintop, 2,000 m, there are few left. However, the muons penetrate far through the earth's atmosphere and many reach the ground.

The muons descend through the earth's atmosphere with a velocity close to c. The minimum time to descend 2,000 m is then

$$T = \frac{2 \times 10^3 \text{ m}}{3 \times 10^8 \text{ m/s}} = 7 \times 10^{-6} \text{ s}.$$

This is more than three times the lifetime; $T/\tau \approx 3$.

The experiment consists of comparing the flux of μ mesons at the top of a mountain with the flux at sea level. We can safely neglect the formation of new mesons in the lower atmosphere or the loss of mesons due to absorption in air. One might expect

$$\frac{\text{flux at sea level}}{\text{flux at mountaintop}} = e^{-T/\tau} = 0.045.$$

[1] B. Rossi and D. B. Hall, *Physical Review,* vol. 59, p. 223, 1941.

[2] An account of the experiment demonstrated in the film is given by D. H. Frisch and J. H. Smith, *American Journal of Physics,* vol. 31, p. 342, 1963.

However, the experimental result disagrees sharply: the ratio is 0.7, corresponding to $T/\tau = 0.3$, which is smaller than the predicted ratio by a factor of 10.

The resolution of the disagreement is that we have neglected time dilation. The lifetime τ refers to the decay of a meson at rest. The μ mesons in the atmosphere are moving at high speed with respect to the laboratories on the mountaintop and at its base. When the muon moves rapidly, the lifetime τ' we observe is increased by time dilation. The observed lifetime is

$$\tau' = \gamma\tau = \frac{\tau}{\sqrt{1 - v^2/c^2}}.$$

To account for the observed muon decay rate, we require $\gamma = 10$. This was found to be the case: by measuring the energy of the mesons, γ was determined, and within experimental error it agreed with the prediction from relativity.

Example 12.6 The Role of Time Dilation in an Atomic Clock

Possibly you have looked through a spectroscope at the light from an atomic discharge lamp. Each line of the spectrum is composed of the light emitted when an atom makes a transition between two of its internal energy states. The lines have different colors because the frequency ν of the light is proportional to the energy change ΔE in the transition. (Atomic spectra are discussed in more detail in Sec. 6.8.) If ΔE is of the order of electron volts, the emitted light is in the optical region ($\nu \approx 10^{15}$ Hz). There are some transitions, however, for which the energy change is so small that the emitted radiation is in the microwave region ($\nu \approx 10^{10}$ Hz). These microwave signals can be detected and amplified electronically. Since the oscillation frequency depends almost entirely on the internal structure of the atom, the signals can serve as a frequency reference to govern the rate of an atomic clock. Atomic clocks are highly stable and relatively immune to external influences.

Each atom radiating at its natural frequency serves as a miniature clock. The atoms are frequently in a gas and move randomly with thermal velocities. Because of their thermal motion, the clocks are not at rest with respect to the laboratory and the observed frequency is shifted by time dilation.

Consider an atom which is radiating its characteristic frequency ν_0 in the rest frame. We can think of the atom's internal harmonic motion as akin to the swinging motion of the pendulum of a grandfather's clock: each cycle corresponds to a complete swing of the pendulum. If the period of the swing is τ_0 seconds in the rest frame, the period in the laboratory is $\tau = \gamma\tau_0$. The observed frequency in the laboratory system is

$$\nu = \frac{1}{\tau} = \frac{1}{\gamma\tau_0} = \frac{\nu_0}{\gamma}$$

$$= \nu_0\sqrt{1 - \frac{v^2}{c^2}}.$$

The shift in the frequency is $\delta\nu = \nu - \nu_0$. If $v^2/c^2 \ll 1$, $\gamma \approx 1 - \frac{1}{2}v^2/c^2$, and the fractional change in frequency is

$$\frac{\delta\nu}{\nu_0} = \frac{\nu - \nu_0}{\nu_0} = -\frac{1}{2}\frac{v^2}{c^2}. \qquad 1$$

A handy way to evaluate the term on the right is to multiply numerator and denominator by M, the mass of the atom:

$$\frac{\delta\nu}{\nu_0} = -\frac{\frac{1}{2}Mv^2}{Mc^2}$$

$\frac{1}{2}Mv^2$ is the kinetic energy due to thermal motion of the atom. This energy increases with the temperature of the gas, and according to an elementary result of statistical mechanics,

$$\tfrac{1}{2}M\overline{v^2} = \tfrac{3}{2}kT,$$

where $\overline{v^2}$ is the average squared velocity, $k = 1.38 \times 10^{-23}$ J/deg is Boltzmann's constant, and T is the absolute temperature.

In the atomic clock known as the hydrogen maser, the reference frequency arises from a transition in atomic hydrogen. M is close to the mass of a proton, 1.67×10^{-27} kg, and using $c = 3 \times 10^8$ m/s, we obtain from Eq. (1),

$$\frac{\delta\nu}{\nu} = -\frac{\frac{3}{2} \times 1.38 \times 10^{-23}}{1.67 \times 10^{-27} \times 9 \times 10^{16}}\,T$$

$$= 1.4 \times 10^{-13}\,T.$$

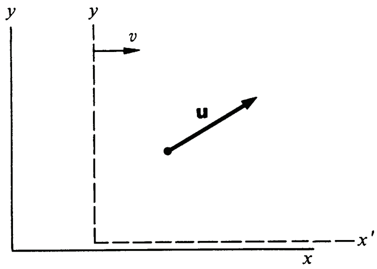
At room temperature, $T = 300$ K (300 degrees on the absolute temperature scale or 27°C), we have

$$\frac{\delta\nu}{\nu} = -4.2 \times 10^{-11}.$$

This, believe it or not, is a sizable effect. In order to correct for time dilation to an accuracy of 1 part in 10^{13}, it is necessary to know the temperature of the radiating atoms to an accuracy of one degree. However, if one wishes to compare frequencies to parts in 10^{15}, the absolute temperature must be known to millidegrees, a much harder task.

12.4 The Relativistic Transformation of Velocity

The starship Enterprise silently glides to the east with speed $0.9c$. At the same time, the starship Fleagle glides to the west with speed $0.9c$. Classically, the relative speed of the ships is $1.8c$, and the Fleagle's crew would see the Enterprise moving away with a speed faster than light. According to special relativity the picture is quite different. To show this we need the relativistic law for the addition of velocities.



Consider a particle with instantaneous velocity $\mathbf{u} = (u_x, u_y)$ in the x, y, z, t system. Our task is to find the corresponding components u'_x, u'_y in the x', y', z', t' system, which moves with speed v along the positive x axis.

From the definition of velocity, we have, in the unprimed system,

$$u_x = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \qquad u_y = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t}.$$

The corresponding components in the primed system are

$$u'_x = \lim_{\Delta t' \to 0} \frac{\Delta x'}{\Delta t'} \qquad u'_y = \lim_{\Delta t' \to 0} \frac{\Delta y'}{\Delta t'}.$$

The problem is to relate displacements and time intervals in the primed system to those in the unprimed system. Using the procedure of Example 12.2 (or simply writing the Lorentz transformations for differentials), we find

$$\Delta x' = \gamma(\Delta x - v\,\Delta t)$$
$$\Delta y' = \Delta y$$
$$\Delta t' = \gamma\left(\Delta t - \frac{v}{c^2}\,\Delta x\right).$$

Hence,

$$\frac{\Delta x'}{\Delta t'} = \frac{\gamma(\Delta x - v\Delta t)}{\gamma[\Delta t - (v/c^2)\Delta x]} = \frac{\Delta x/\Delta t - v}{1 - (v/c^2)(\Delta x/\Delta t)}.$$

Next we take the limit $\Delta t \rightarrow 0$. Since $\Delta x = u_x\, \Delta t$, $\Delta x \rightarrow 0$ when $\Delta t \rightarrow 0$. The Lorentz transformations show that $\Delta x'$ and $\Delta t'$ also approach zero. Using $u_x' = \lim_{\Delta t' \rightarrow 0} (\Delta x'/\Delta t')$, we obtain

$$u_x' = \frac{u_x - v}{1 - vu_x/c^2}. \tag{12.2a}$$

Similarly,

$$u_y' = \frac{u_y}{\gamma[1 - vu_x/c^2]}. \tag{12.2b}$$

By symmetry, u_z' behaves like u_y':

$$u_z' = \frac{u_z}{\gamma[1 - vu_x/c^2]}. \tag{12.2c}$$

These transformations can be inverted by changing the sign of v:

$$u_x = \frac{u_x' + v}{1 + vu_x'/c^2} \tag{12.3a}$$

$$u_y = \frac{u_y'}{\gamma[1 + vu_x'/c^2]} \tag{12.3b}$$

$$u_z = \frac{u_z'}{\gamma[1 + vu_x'/c^2]}. \tag{12.3c}$$

In these formulas, $\gamma = 1/\sqrt{1 - v^2/c^2}$ as before.

Equation (12.2a) or (12.3a) is the relativistic law for the addition of velocities. For $v \ll c$, we obtain the Galilean result $u_x' = u_x - v$.

Returning to the problem of the two starships, let $u_x = 0.9c$ be the speed of the Enterprise relative to the ground, and $v = -0.9c$ be the speed of the Fleagle relative to the ground. The velocity of the Enterprise relative to the Fleagle is, from Eq. (12.2a),

$$\begin{aligned} u_x' &= \frac{0.9c - (-0.9c)}{1 - [(-0.9c)(0.9c)]} \\ &= \frac{1.8c}{1.81} \\ &= 0.99c. \end{aligned}$$

The relative speed is less than c. The relativistic transformation of velocities assures that we cannot exceed the velocity of light by changing reference frames.

The limiting case is $u_x = c$. The velocity in the moving system is then

$$u_x' = \frac{c - v}{1 - vc/c^2} = c,$$

independent of v. This agrees with the postulate we originally built into the Lorentz transformations: the velocity of light is the same for all observers. Furthermore, it suggests that the velocity of light plays the role of an ultimate speed in the theory of relativity.

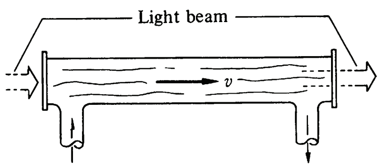
Example 12.7 The Speed of Light in a Moving Medium

As an exercise in the relativistic addition of velocities, let us find how the motion of a medium, such as water, influences the speed of light.

The velocity of light in matter is less than c. The index of refraction, n, is used to specify the speed in a medium:

$$n = \frac{c}{\text{velocity of light in the medium}}$$

$n = 1$ corresponds to empty space; in matter $n > 1$. The slowing can be appreciable: for water $n = 1.3$.



The problem is to find the speed of light through a moving liquid. For instance, consider a tube filled with water. If the water is at rest, the velocity of light in the water with respect to the laboratory is $u = c/n$. What is the speed of light when the water is flowing with speed v?

Consider the speed of light in water as observed in a coordinate system x', y' moving with the water. The speed is

$$u' = \frac{c}{n}.$$

The speed in the laboratory is, by Eq. (12.3a),

$$u = \frac{u' + v}{1 + u'v/c^2} = \frac{c/n + v}{1 + v/nc} = \frac{c}{n}\left(\frac{1 + nv/c}{1 + v/nc}\right).$$

If we expand the last term and neglect terms of order $(v/c)^2$ and smaller, we obtain

$$u = \frac{c}{n}\left(1 + \frac{nv}{c} - \frac{v}{nc}\right) = \frac{c}{n} + v\left(1 - \frac{1}{n^2}\right).$$

The light appears to be dragged by the fluid, but not completely. Only the fraction $f = 1 - 1/n^2$ of the fluid velocity is added to the speed of light c/n. This effect was observed experimentally in 1851 by Fizeau, although it was not explained satisfactorily until the advent of relativity.

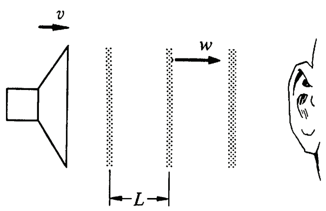
12.5 The Doppler Effect

The roar of a car or motorcycle zooming past is characterized by a rapid drop in pitch as the vehicle goes by. The effect is quite noticeable if you listen for it at the side of a road. (It is the sound most people make when trying to mimic a near miss by a speeding car.) The decrease in frequency of all the sounds from the car as it goes by is due to the Doppler effect. In general, the Doppler effect is a shift in frequency due to the motion of a source or an observer. The Doppler shift occurs for light as well as sound. Our knowledge of the motion of distant receding galaxies comes from studies of the Doppler shift of their spectral lines. More prosaic applications of the Doppler effect include satellite tracking and radar speed traps.

We shall start by examining the Doppler shift in sound—a situation we can treat classically.

The Doppler Shift in Sound

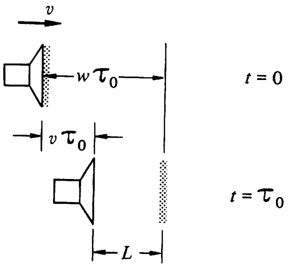
Sound travels through a medium, such as air, with a speed w determined by the properties of the medium, independent of the motion of the source.



Consider a source of sound which is moving with velocity v through the medium toward an observer at rest. To simplify the geometry we shall restrict ourselves for the present to the case where the observer is along the line of motion. We can regard the sound as a regular series of pulses separated by time $\tau_0 = 1/\nu_0$, where ν_0 is the number of pulses per second generated by the source. (ν_0 corresponds to the frequency of sound from the source.) The situation is shown in the sketch.

In time T the sound travels a distance wT, and if the pulses are separated by distance L, the number reaching the observer is wT/L. The rate at which the pulses arrive is w/L, and this is the frequency of sound ν_D heard by the observer:

$$\nu_D = \frac{w}{L}.$$



To determine L, consider a pulse emitted at $t = 0$ and the next pulse emitted at $t = \tau_0$. During the interval τ_0 the first pulse travels distance $w\tau_0$ in the medium, and the source travels distance $v\tau_0$. The distance between the pulses is therefore

$$L = w\tau_0 - v\tau_0$$

$$= (w - v)\frac{1}{\nu_0}.$$

Hence,

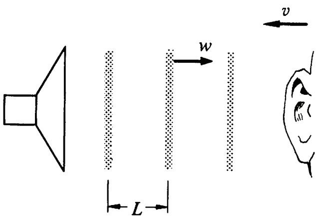
$$\nu_D = \frac{w}{L}$$

$$= \nu_0 \frac{w}{w - v}$$

or

$$\nu_D = \nu_0 \frac{1}{1 - (v/w)}. \qquad \text{(Moving source.)} \qquad 12.4$$

For an approaching source, v is positive and $\nu_D > \nu_0$. For a receding source, v is negative and $\nu_D < \nu_0$. Qualitatively, this accounts for the drop in pitch of the sound of a car as it goes by.



The situation is somewhat different if the source is at rest in the medium and the observer is moving with speed v toward the source. The situation is shown in the sketch. The speed of the pulses relative to the observer is $w + v$. The rate at which pulses arrive is

$$\nu_D = \frac{w + v}{L}.$$

Since the source is at rest, $L = w\tau_0 = w/\nu_0$, and

$$\nu_D = \nu_0 \frac{w + v}{w} = \nu_0\left(1 + \frac{v}{w}\right). \qquad \text{(Moving observer.)} \qquad 12.5$$

This differs from the result for a moving source, Eq. (12.4), although the results agree to order v/w. The situation is not symmetric; if ν_0, v, and w are known, we can tell whether it is the observer or the source which is moving by measuring ν_D carefully. The reason is that in the case of sound there is a medium, the air, to which motion can be referred.

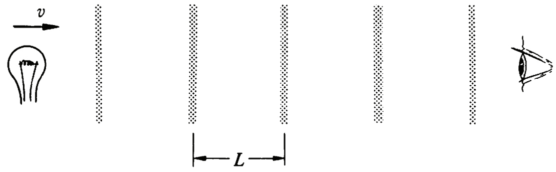
If it were possible to apply these results to light waves in space, we would be able to distinguish which of two inertial systems was at rest. This would contradict the principle of special relativity that only the relative motion of inertial systems is observable. To resolve this difficulty, we turn now to a relativistic derivation of the Doppler effect.

Relativistic Doppler Effect

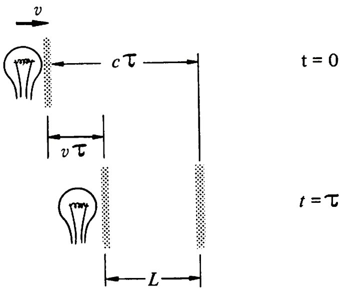
A light source flashes with period $\tau_0 = 1/\nu_0$ in its rest frame. The source is moving toward an observer with velocity v. Due to time dilation, the period in the observer's rest frame is

$$\tau = \gamma\tau_0.$$

Since the speed of light is a universal constant, the pulses arrive at the observer with speed c. It is for this reason that the relative velocity alone plays a role in the Doppler effect for light. In the classical case, the pulses arrive with a speed dependent on the state of motion of the observer relative to the medium.



The frequency of the pulses is $\nu_D = c/L$, where L is the separation in the observer's frame. Since the source is moving toward the observer,



$$L = c\tau - v\tau = (c - v)\tau$$

and

$$\nu_D = \frac{c}{(c - v)\tau}$$

$$= \frac{1}{1 - v/c}\frac{1}{\gamma\tau_0}$$

or

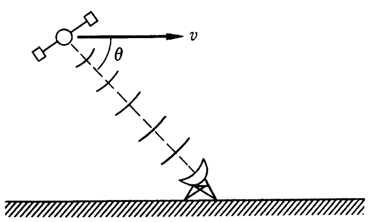
$$\nu_D = \nu_0\frac{\sqrt{1 - v^2/c^2}}{1 - v/c}.$$

This reduces to

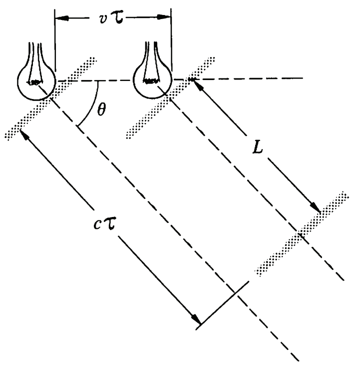
$$\nu_D = \nu_0\sqrt{\frac{1 + v/c}{1 - v/c}}. \qquad 12.6$$

ν_D is the frequency in the observer's rest frame and v is the relative speed of source and observer. As we expect, there is no mention of motion relative to a medium. The relativistic result plays no favorites with the classical results; it disagrees with both and, in fact, turns out to be their geometric mean.

The Doppler Effect for an Observer off the Line of Motion



So far we have restricted ourselves to the Doppler effect for a source and observer along the line of motion. However, consider a satellite broadcasting a radio beacon signal to a ground tracking station which monitors the Doppler shifted frequency. Although our earlier results do not apply to such a case, we can readily generalize the method to find the Doppler effect when the observer is at angle θ from the line of motion. We shall again visualize the source as a flashing light. The period of the flashes in the observer's rest frame is $\tau = \gamma\tau_0$, as before. The frequency seen by the observer is c/L. Since the source moves distance $v\tau$ between flashes, it is apparent from the lower sketch that



$$L = c\tau - v\tau\cos\theta$$
$$= (c - v\cos\theta)\tau.$$

(We assume that the source and observer are so far apart that θ is effectively constant between pulses.) Hence

$$\nu_D = \frac{c}{L}$$
$$= \frac{c}{(c - v\cos\theta)\tau_0\gamma}$$

or

$$\nu_D = \nu_0\frac{\sqrt{1 - v^2/c^2}}{1 - (v/c)\cos\theta}. \qquad 12.7$$

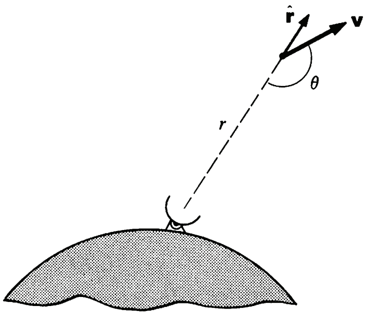
In this result, θ is the angle measured in the rest frame of the observer. Along the line of motion, $\theta = 0$ and we recover our previous result for that case, Eq. (12.6). At $\theta = \pi/2$ the relative velocity between source and observer is zero. However, even in this case there is a shift in frequency; ν_D differs from ν_0 by the factor $\sqrt{1 - v^2/c^2}$. This "transverse" Doppler effect is due to time dilation. The flashing lamp is effectively a moving clock.

The relativistic Doppler effect agrees with the classical result to order v/c, so that any experiment to differentiate between them must be sensitive to effects of order $(v/c)^2$, a difficult task.

The relativistic expression was confirmed by Ives and Stilwell in 1938 by observations on the spectral light from fast moving atoms.

One of the more interesting practical applications of the Doppler effect is in navigational systems, as the following example explains.

Example 12.8 Doppler Navigation



The Doppler effect can be used to track a moving body, such as a satellite, from a reference point on the earth. The method is remarkably accurate; changes in the position of a satellite 10^8 m away can be determined to a fraction of a centimeter.

Consider a satellite moving with velocity $\mathbf{v}$ at some distance r from a ground station. An oscillator on the satellite broadcasts a signal with proper frequency ν_0. Since $v \ll c$ for satellites, we can approximate Eq. (12.7) by retaining only terms of order v/c. Then the frequency ν_D received by the ground station can be written

$$\nu_D \approx \frac{\nu_0}{1 - (v/c)\cos\theta}$$
$$\approx \nu_0\left(1 + \frac{v}{c}\cos\theta\right).$$

There is an oscillator in the ground station identical to the one in the satellite, and by simple electronic methods the difference frequency ("beat" frequency) $\nu_D - \nu_0$ can be measured:

$$\nu_D - \nu_0 = \nu_0 \frac{v}{c}\cos\theta.$$

The radial velocity of the satellite is

$$\frac{dr}{dt} = \hat{\mathbf{r}}\cdot\mathbf{v}$$
$$= -v\cos\theta.$$

Hence

$$\frac{dr}{dt} = -\frac{c}{\nu_0}(\nu_D - \nu_0)$$
$$= -\lambda_0(\nu_D - \nu_0),$$

where $\lambda_0 = c/\nu_0$ is the wavelength of the radiation.

ν_D varies in time as the satellite's velocity and direction change. To find the total radial distance traveled between times T_a and T_b, we integrate the above expression with respect to time:

$$\int_{T_a}^{T_b}\left(\frac{dr}{dt}\right)dt = -\lambda_0\int_{T_a}^{T_b}(\nu_D - \nu_0)\,dt$$
$$r_b - r_a = -\lambda_0\int_{T_a}^{T_b}(\nu_D - \nu_0)\,dt.$$

The integral is the number of cycles N_{ba} of the beat frequency which occur in the interval T_a to T_b. (One cycle occurs in a time $\tau = 1/(\nu_D - \nu_0)$, so that $\int dt/\tau$ is the total number of cycles.) Hence

$$r_b - r_a = -\lambda_0 N_{ba}.$$

This result has a simple interpretation: whenever the radial distance increases by one wavelength, the phase of the beat signal decreases one cycle. Similarly, when the radial distance decreases one wavelength, the phase of the beat signal increases by one cycle.

Satellite communication systems operate at a typical wavelength of 10 cm, and since the beat signal can be measured to a fraction of a cycle, satellites can be tracked to about 1 cm. If the satellite and ground-based oscillators do not each stay tuned to the same frequency, ν_0, there will be an error in the beat frequency. To avoid this problem a two-way Doppler tracking system can be used in which a signal from the ground is broadcast to the satellite which then amplifies it and relays it back to the ground. This has the added advantage of doubling the Doppler shift, increasing the resolution by a factor of 2.

We sketched the principles of Doppler navigation for the classical case $v \ll c$. For certain tracking applications the precision is so high that relativistic effects must be taken into account.

As we have already shown, a Doppler tracking system also gives the instantaneous radial velocity of the satellite $v_r = -c(\nu_D - \nu_0)/\nu_0$. This is particularly handy, since both velocity and position are needed to check satellite trajectories. A more prosaic use of this result is in police radar speed monitors: a microwave signal is reflected from an oncoming car and the beat frequency of the reflected signal reveals the car's speed.

12.6 The Twin Paradox

The kinematical effects we have analyzed in this chapter depend on the *relative* velocity of two systems; such phenomena as Lorentz contraction, time dilation, and the Doppler shift give no clue as to which of two systems is at rest and which is moving, nor can they do so within the framework of relativity, which postulates that all inertial systems are equivalent. There is no such equivalence between *non*inertial systems. Indeed, there is little difficulty in deciding whether or not an isolated system is accelerating.

Failure to appreciate this point was responsible for a vociferous controversy over the so-called "twin paradox." The problem is of interest because it affords a good illustration of the physical difference between inertial and noninertial systems.

The paradox is as follows: two identical twins, Castor and Pollux, A and B for short, have identical clocks. B sets out on a long space voyage while A remains home. A constantly observes B's clock and sees that it is running slow due to time dilation. Eventually B returns home. Since B's clock has run slow throughout the trip, A concludes that B is younger than A at the end of the journey. But suppose we look at the situation from B's point of view. Since time dilation depends only on relative motion, during the trip B sees A's clock running slow, and when the trip is finished B concludes that A is younger than B. Obviously both twins can't be right. Is either twin really younger?

The explanation lies in the fact that the situation is *not* equivalent from the point of view of each twin. A's system is inertial throughout, but B must change his velocity at some time in order to return to the starting point. While the velocity is changing, B's system is not inertial. There is no doubt as to which twin is really accelerating. If each were carrying an accelerometer, such as a mass on a spring, A's would remain at zero while B's would show a large deflection at the turning point. It is apparent that the systems are not equivalent.

We cannot apply special relativity to determine the coordinates of events in noninertial frames. Fortunately, it is possible to determine what B will observe during turnaround by introducing the idea of the Doppler shift.

To make the argument quantitative, suppose that the relative velocity is v. A observes that B travels away a distance L in time $T = L/v$. B then rapidly reverses his motion and returns with the same velocity. The time for the return trip is also T. We shall neglect the time it takes B to reverse his motion since if T is sufficiently long, the turnaround time is negligible. (Nothing anomalous happens to B's clock during turnaround; A simply observes a varying dilation factor while the velocity is changing.)

Neglecting this small turnaround correction, A observes a total elapsed time T'_B on B's moving clock which is related to the time on A's own clock $T_A = 2T$ by

$$\begin{aligned} T'_B &= \frac{T_A}{\gamma} \\ &= T_A\sqrt{1 - \frac{v^2}{c^2}}. \end{aligned} \qquad 12.8$$

A concludes that B is younger.

$$\frac{\text{aging of } B}{\text{aging of } A} = \frac{T'_B}{T_A} = \sqrt{1 - \frac{v^2}{c^2}} \qquad \text{(As viewed by } A.\text{)} \qquad 12.9$$

Now let us look at the situation from B's point of view. Except for the turnaround time, B's observations are similar to A's. B sees A go away for distance L with velocity $-v$ and return. This takes time $T_B = 2T$ on B's clock, and if B sees time T'_A elapse on A's clock, then

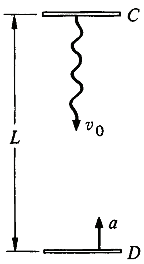
$$\begin{aligned} T'_A &= \frac{T_B}{\gamma} \\ &= T_B\sqrt{1-\frac{v^2}{c^2}}. \end{aligned} \qquad 12.10$$

B seems to conclude that A is younger.

$$\frac{\text{aging of } B}{\text{aging of } A} \stackrel{?}{=} \frac{T_B}{T'_A} = \frac{1}{\sqrt{1-v^2/c^2}}. \qquad \text{(As viewed by } B.\text{)} \qquad 12.11$$

This is the paradox: A thinks that B is younger and B thinks that A is younger.

Now consider what happens to B during turnaround. He experiences an acceleration as if he were in a gravitational field. According to the discussion of the principle of equivalence in Chap. 8, clocks run at different rates in a gravitational field—this is the origin of the gravitational red shift. For this reason, B sees A's clock run fast during turnaround and, as we shall show, this puts A's clock ahead. However, instead of involving the gravitational red shift, we shall derive the result from simple kinematics.



Consider a clock C which has period τ_0 in its rest frame and which emits signals at frequency $\nu_0 = 1/\tau_0$. An observer D is at rest a distance L away and starts accelerating toward C at rate a when the signal of frequency ν_0 leaves C. The signal arrives at time $t_0 \approx L/c$. (We assume that D has not moved appreciably in time t_0, and that his velocity is so low that relativistic effects are negligible.) When the signal arrives, D is moving toward C at velocity $v = at_0 = aL/c$ and the observed frequency, ν', is Doppler shifted. From Eq. (12.6) we have

$$\begin{aligned} \nu' &= \nu_0\sqrt{\frac{1+v/c}{1-v/c}} \\ &\approx \nu_0\left(1+\frac{v}{c}\right) \\ &= \nu_0\left(1+\frac{aL}{c^2}\right), \end{aligned}$$

where we have neglected terms of order $(v/c)^2$. Since $\nu' > \nu_0$, C's clock appears to run faster than if there were no acceleration. If D's clock records a time interval

$$T_D = 1/\nu',$$

then C's clock marks off an interval

$$T_C = 1/\nu_0.$$

Hence,

$$\begin{aligned} T_C &= T_D \frac{\nu'}{\nu_0} \\ &= T_D\left(1 + \frac{aL}{c^2}\right). \end{aligned}$$

Applying this to the twins, suppose that B accelerates uniformly at rate a toward A during turnaround. B notes on his own clock that the turnaround time is T_t. He notes that A's clock marks off an interval

$$T'_t = T_t\left(1 + \frac{aL}{c^2}\right).$$

Since the velocity changes by $2v$ during turnaround, $T_t = 2v/a$. Therefore,

$$\begin{aligned} T'_t &= T_t + \frac{2v}{a}\frac{aL}{c^2} \\ &= T_t + \frac{2vL}{c^2}. \end{aligned}$$

The total length of the trip is $2L = vT_B$. Hence, the total time that B observes on A's clock during turnaround is

$$T'_t = T_t + \frac{v^2}{c^2}T_B.$$

The total time that B observes on A's clock during the entire trip is

$$\begin{aligned} (T'_A)_{\text{total}} &= T'_A + T_t + \frac{v^2}{c^2}T_B \\ &= T_B\sqrt{1 - \frac{v^2}{c^2}} + T_t + \frac{v^2}{c^2}T_B, \end{aligned}$$

where we have used $T'_A = T_B/\gamma$, Eq. (12.10). We shall again neglect the turnaround time. The Doppler shift correction during turnaround is valid to order v^2/c^2 and to this approximation,

$$(T'_A)_{\text{total}} = T_B\left(1 - \frac{1}{2}\frac{v^2}{c^2} + \frac{v^2}{c^2}\right)$$
$$= T_B\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right).$$

The result of this argument is that from B's point of view,

$$\frac{\text{aging of } B}{\text{aging of } A} = \frac{T_B}{(T'_A)_{\text{total}}} = \frac{1}{1 + \frac{1}{2}v^2/c^2} \approx 1 - \frac{1}{2}\frac{v^2}{c^2}.$$

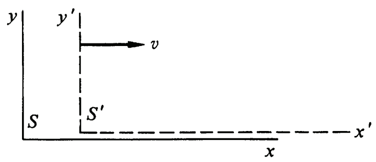
We have already shown, Eq. (12.9), that from A's point of view

$$\frac{\text{aging of } B}{\text{aging of } A} = \frac{T'_B}{T_A} = \sqrt{1 - \frac{v^2}{c^2}} \approx 1 - \frac{1}{2}\frac{v^2}{c^2}.$$

The formerly identical twins are in agreement; A has aged more than B. The paradox is resolved.

Our analysis is valid only to order v^2/c^2. To this order, the special theory of relativity led to no contradictions as long as we treated the accelerated reference frame separately. An exact calculation appears to require the general theory of relativity.

Problems



In these problems S refers to an inertial system x, y, z, t and S' refers to an inertial system x', y', z', t', moving along the x axis with speed v relative to S. The origins coincide at $t = t' = 0$. Take $c = 3 \times 10^8$ m/s.

12.1 Assume that $v = 0.6c$. Find the coordinates in S' of the following events.

a. $x = 4$ m, $t = 0$ s.

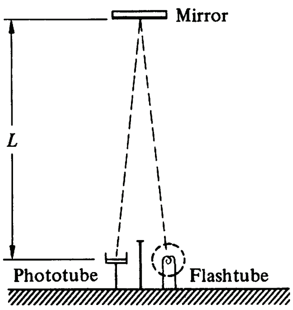
b. $x = 4$ m, $t = 1$ s.

c. $x = 1.8 \times 10^8$ m, $t = 1$ s.

d. $x = 10^9$ m, $t = 2$ s.

12.2 An event occurs in S at $x = 6 \times 10^8$ m, and in S' at $x' = 6 \times 10^8$ m, $t' = 4$ s. Find the relative velocity of the systems.

12.3 The clock in the sketch on the opposite page can provide an intuitive explanation of the time dilation formula. The clock consists of a flash tube, mirror, and phototube. The flash tube emits a pulse of light which



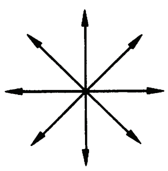
travels distance L to the mirror and is reflected to the phototube. Every time a pulse hits the phototube it triggers the flash tube. Neglecting time delay in the triggering circuits, the period of the clock is $\tau_0 = 2L/c$.

Now examine the clock in a coordinate system moving to the left with uniform velocity v. In this system the clock appears to move to the right with velocity v. Find the period of the clock in the moving system by direct calculation, using only the assumptions that c is a universal constant, and that distance perpendicular to the line of motion is unaffected by the motion. The result should be identical to that given by the Lorentz transformations: $\tau = \tau_0/\sqrt{1 - v^2/c^2}$.

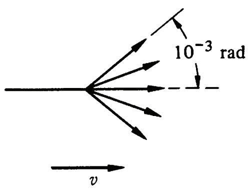
12.4 A light beam is emitted at angle θ_0 with respect to the x' axis in S'.

a. Find the angle θ the beam makes with respect to the x axis in S.

Ans. $\cos\theta = (\cos\theta_0 + v/c)/(1 + v/c\cos\theta_0)$



Rest frame



b. A source which radiates light uniformly in all directions in its rest frame radiates strongly in the forward direction in a frame in which it is moving with speed v close to c. This is called the headlight effect; it is very pronounced in synchrotrons in which electrons moving at relativistic speeds emit light in a narrow cone in the forward direction. Using the result of part *a*, find the speed of a source for which half the radiation is emitted in a cone subtending 10^{-3} rad.

Ans. $v = c(1 - 5 \times 10^{-7})$

12.5 An observer sees two spaceships flying apart with speed $0.99c$. What is the speed of one spaceship as viewed by the other?

Ans. $0.99995c$

12.6 A rod of proper length l_0 oriented parallel to the x axis moves with speed u along the x axis in S. What is the length measured by an observer in S'?

Ans. $l = l_0[(c^2 - v^2)(c^2 - u^2)]^{\frac{1}{2}}/(c^2 - uv)$

12.7 One of the most prominent spectral lines of hydrogen is the H_α line, a bright red line with a wavelength of 656.1×10^{-9} m.

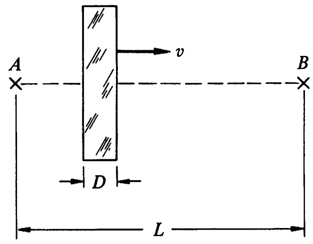
a. What is the expected wavelength of the H_α line from a star receding with a speed of 3,000 km/s?

Ans. 662.7×10^{-9} m

b. The H_α line measured on earth from opposite ends of the sun's equator differ in wavelength by 9×10^{-12} m. Assuming that the effect is caused by rotation of the sun, find the period of rotation. The diameter of the sun is 1.4×10^6 km.

Ans. 25 d

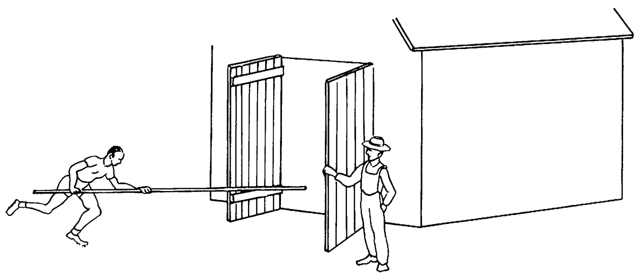
12.8 The frequency of light reflected from a moving mirror undergoes a Doppler shift because of the motion of the image. Find the Doppler shift of light reflected directly back from a mirror which is approaching the observer with speed v, and show that it is the same as if the image were moving toward the observer at speed $2v/(1 + v^2/c^2)$.



12.9 A slab of glass moves to the right with speed v. A flash of light is emitted from A and passes through the glass to arrive at B, a distance L away. The glass has thickness D in its rest frame, and the speed of light in the glass is c/n. How long does it take the light to go from A to B?

Ans. clue. If $v = 0$, $T = [L + (n - 1)D]/c$; if $v = c$, $T = L/c$

12.10 Here is the pole-vaulter paradox. A pole-vaulter and a farmer have the following bet: the pole-vaulter has a pole of length l_0, and the farmer has a barn $\frac{3}{4}l_0$ long. The farmer bets that he can shut the door of the barn with the pole completely inside. The bet being made, the farmer asks the pole-vaulter to run into the barn with a speed of $v = c\sqrt{3}/2$. In this case the farmer observes the pole to be Lorentz contracted to $l = l_0/2$, and the pole fits into the barn with ease. He slams the door the instant the pole is inside, and claims the bet. The pole-vaulter disagrees: he sees the barn contracted by a factor of 2, and so the pole can't possibly fit inside. How would you settle the disagreement? Is the Lorentz contraction "real" in this problem? (*Hint*: Consider events at the ends of the pole from the point of view of each observer.)



12.11 The relativistic transformation of acceleration from S' to S can be found by extending the procedure of Sec. 12.4. The most useful transformation is for the case in which the particle is instantaneously at rest in S' but is accelerating at rate a_0 in S', parallel to the x' axis.

Show that for this case the x acceleration in S is given by $a_x = a_0/\gamma^3$.

12.12 The relativistic transformation for acceleration derived in the last problem shows the impossibility of accelerating a system to a velocity greater than c. Consider a rocketship which accelerates at constant rate a_0 as measured by an accelerometer carried aboard, for instance a a mass stretching a spring.

a. Find the speed after time t for an observer in the system in which the rocketship was originally at rest.

Ans. $v = a_0t/\gamma$, or $v = a_0t/\sqrt{1 + (a_0t/c)^2}$

b. The speed predicted classically is $v_0 = a_0t$. What is the actual speed for the following cases: $v_0 = 10^{-3}c$, c, 10^3c.

Ans. $v = v_0(1 - 5 \times 10^{-7})$, $c/\sqrt{2}$, $c(1 - 5 \times 10^{-7})$

12.13 A young man voyages to the nearest star, α Centauri, 4.3 light-years away. He travels in a spaceship at a velocity of $c/5$. When he returns to earth, how much younger is he than his twin brother who stayed home?

12.14 Any quantity which is left unchanged by the Lorentz transformations is called a *Lorentz invariant*. Show that Δs is a Lorentz invariant, where

$$\Delta s^2 = (c\,\Delta t)^2 - (\Delta x^2 + \Delta y^2 + \Delta z^2).$$

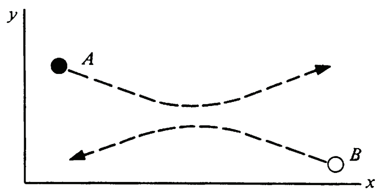
Here Δt is the interval between two events and $(\Delta x^2 + \Delta y^2 + \Delta x^2)^{\frac{1}{2}}$ is the distance between them in the same inertial system.

13 RELATIVISTIC MOMENTUM AND ENERGY

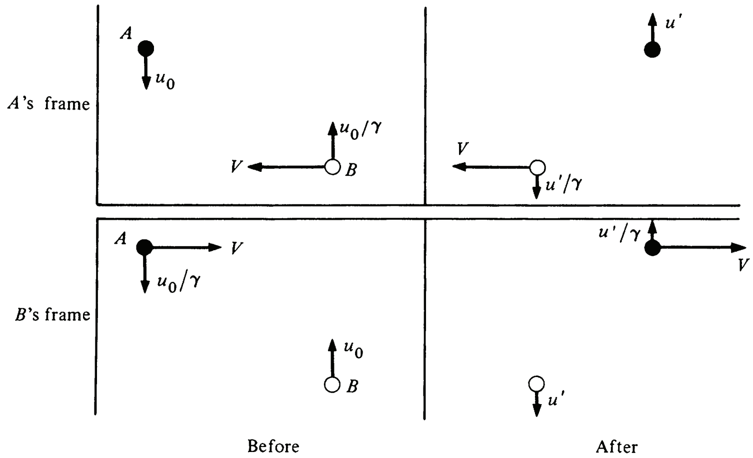
13.1 Momentum

In the last chapter we saw how the postulates of special relativity lead in a natural way to kinematical relations which agree with newtonian relations at low velocity but depart markedly for velocities approaching c. We turn now to the problem of investigating the implications of special relativity for dynamics. One approach would be to develop a formal procedure for writing the laws of physics in a form which satisfies the postulates of special relativity. Such a procedure is actually possible; it involves the concepts of four-vectors and relativistic invariance, and we shall pursue it in the next chapter. However, here we shall take another approach, one which is not as powerful or as economical as the method of four-vectors, but which has the advantage of using physical arguments to show the relation between the familiar concepts of classical mechanics and their relativistic counterparts.

First we shall focus on conservation of momentum and find what modifications are needed to preserve this principle in relativistic mechanics. This is a technique often used in extending the frontiers of physics: by reformulating conservation laws so that they are preserved in new situations, we are quite naturally led to generalizations of familiar concepts. In particular, as the following argument shows, we must modify our idea of mass to preserve conservation of momentum under relativistic transformations.



Consider a glancing elastic collision between two identical particles, A and B. We are going to view the collision in two special frames: A's frame, the frame moving along the x axis with A, and B's frame, the frame moving along the x axis with B. We



take the collisions to be completely symmetrical. Each particle has the same y speed u_0 in its own frame before the collision, as shown in the sketches. The effect of the collision is to alter the y velocities but leave the x velocities unchanged.

The relative x velocity of the frames is V and by the law of transformation of velocities, Eq. (12.2), the y velocity of the opposite particle in each frame is $u_0/\gamma = u_0\sqrt{1 - V^2/c^2}$.

After the collisions the y velocities have reversed their directions as shown in the sketches. The situation remains symmetrical. If the y speed of A and B in their own frames is u', the y speed of the other particle is u'/γ.

Our task is to find a conserved quantity analogous to classical momentum. We suppose that the momentum of a particle moving with velocity $\mathbf{w}$ is

$$\mathbf{p} = m(w)\mathbf{w},$$

where $m(w)$ is a scalar quantity, yet to be determined, analogous to newtonian mass, but one which may depend on the speed w.

The x momentum in A's frame is due entirely to particle B. Before the collision B's speed is $w = (V^2 + u_0^2/\gamma^2)^{\frac{1}{2}}$, and after the collision it is $w' = (V^2 + u'^2/\gamma^2)^{\frac{1}{2}}$. Imposing conservation of momentum in the x direction yields

$$m(w)V = m(w')V.$$

It follows that $w = w'$, so that

$$u' = u_0.$$

Next we write the statement of the conservation of momentum in the y direction, as evaluated in A's frame. Equating the y momentum before and after the collision gives

$$-m(u_0)u_0 + m(w)\frac{u_0}{\gamma} = m(u_0)u_0 - m(w)\frac{u_0}{\gamma}$$

or

$$m(w) = \gamma m(u_0).$$

In the limit $u_0 \to 0$, $m(u_0) \to m(0)$, which we take to be the newtonian mass, or "rest mass" m_0 of the particle. In this limit, $w = V$. Hence

$$m(V) = \gamma m(0) = \frac{m_0}{\sqrt{1 - V^2/c^2}}. \qquad 13.1$$

We have found the dependence of m on speed. In general, therefore,

$$\mathbf{p} = \frac{m_0\mathbf{u}}{\sqrt{1 - u^2/c^2}} = m\mathbf{u}$$

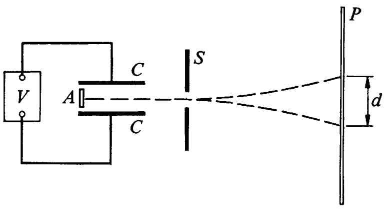
for a particle moving with arbitrary velocity $\mathbf{u}$, where

$$m = \frac{m_0}{\sqrt{1 - u^2/c^2}}. \tag{13.2}$$

Example 13.1 Velocity Dependence of the Electron's Mass

At the beginning of the twentieth century there were several speculative theories which predicted that the mass of an electron varies with its speed. These theories were based on various models of the structure of the electron. The principal theories were those of Abraham (1902), which predicted $m = m_0[1 + \frac{2}{5}(v^2/c^2)]$ for $v \ll c$,† and of Lorentz (1904), which gave $m = m_0/\sqrt{1 - v^2/c^2} \approx m_0[1 + \frac{1}{2}(v^2/c^2)]$. The Abraham theory, which retained the idea of the ether drift and absolute motion, predicted no time dilation effect. Lorentz' result, while identical in form to that published by Einstein in 1905, was derived using the ad hoc Lorentz contraction and did not possess the generality of Einstein's theory.

Experimental work on the effect of velocity on the electron's mass was initiated by Kaufmann in Göttingen in 1902. His data favored the theory of Abraham, and in a 1906 paper he rejected the Lorentz-Einstein results. However, further work by Bestelmeyer (1907) in Göttingen and Bucherer (1909) in Bonn revealed errors in Kaufmann's work and confirmed the Lorentz-Einstein formula.



Physicists were in agreement that the force on a moving electron in an applied electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ is $q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ (the units are SI), where q is the electron's charge and $\mathbf{v}$ its velocity. Bucherer employed this force law in the apparatus shown at left. The apparatus is evacuated and immersed in an external magnetic field $\mathbf{B}$ perpendicular to the plane of the sketch. The source of the electrons A is a button of radioactive material, generally radium salts. The emitted electrons ("beta rays") have a broad energy spectrum extending to 1 MeV or so. To select a single speed, the electrons are passed through a "velocity filter" composed of a transverse electric field $\mathbf{E}$ (produced between two parallel metal plates C by the battery V) together with the magnetic field $\mathbf{B}$. $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{v}$ are mutually perpendicular. The transverse force is

† Abraham's full result was

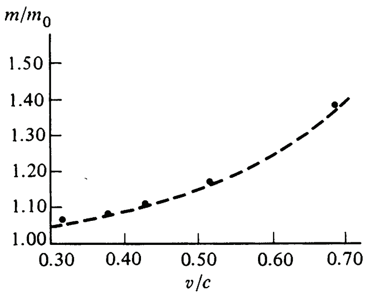
$$m = m_0\frac{3}{4}\frac{1}{\beta^2}\left[\left(\frac{1+\beta^2}{2\beta}\right)\ln\left(\frac{1+\beta}{1-\beta}\right) - 1\right],$$

where $\beta = v/c$.

zero when $qE = qvB$, so that electrons with $v = E/B$ are undeflected and are able to pass through the slit S.

Beyond S only the magnetic field acts. The electrons move with constant speed v and are bent into a circular path by the magnetic force $q\mathbf{v} \times \mathbf{B}$. The radius of curvature R is given by $mv^2/R = qvB$, or $R = mv/qB = (m/q)(E/B^2)$.

The electrons eventually strike the photographic plate P, leaving a trace. By reversing $\mathbf{E}$ and $\mathbf{B}$, the sense of deflection is reversed. R is found from a measurement of the total deflection d and the known geometry of the apparatus. E and B are found by standard techniques. By finding R for different velocities, the velocity dependence of m/q can be studied. We believe that charge does not vary with velocity (otherwise an atom would not stay strictly neutral in spite of how the energy of its electrons varied), so that the variation of m/q can be attributed to variation in m alone.



The graph shows Bucherer's data together with a dashed line corresponding to the Einstein prediction $m = m_0/\sqrt{1 - v^2/c^2}$. The agreement is striking.

The velocity filter with crossed $\mathbf{E}$ and $\mathbf{B}$ fields was used by Bestelmeyer and by Bucherer. (Bucherer attributes the design to J. J. Thomson, discoverer of the electron.) Kaufmann, on the other hand, used transverse $\mathbf{E}$ and $\mathbf{B}$ fields which were parallel to one another, and this probably caused his erroneous results. His configuration did not select velocities; instead, all the electrons were spread into a two dimensional trace on the photographic plate. Electrons of different speeds followed different deflected paths between the plates C, and nonuniformity of the $\mathbf{E}$ field gave rise to substantial errors.

In recent years the relativistic equations of motion have been used to design high energy electron and proton accelerators. For protons, accelerators have been operated with m/m_0 up to 200, while for electrons the ratio $m/m_0 = 40{,}000$ has been reached. The successful operation of these machines leaves no doubt that the relativistic results are accurate.

13.2 Energy

By generalizing the classical concept of energy, we can find a corresponding relativistic quantity which is also conserved. From the discussion in Chap. 4 we can write the kinetic energy of a particle, K, as

$$K_b - K_a = \int_a^b \frac{d\mathbf{p}}{dt} \cdot d\mathbf{r}.$$

For a classical particle moving with velocity **u**, $\mathbf{p} = m\mathbf{u}$, where m is constant. Then

$$K_b - K_a = \int_a^b \frac{d}{dt}(m\mathbf{u}) \cdot d\mathbf{r}$$
$$= \int_a^b m\frac{d\mathbf{u}}{dt} \cdot \mathbf{u}\, dt$$
$$= \int_a^b m\mathbf{u} \cdot d\mathbf{u}.$$

Using the identity $\mathbf{u} \cdot d\mathbf{u} = \frac{1}{2}d(\mathbf{u} \cdot \mathbf{u}) = \frac{1}{2}d(u^2) = u\,du$, we obtain $K_b - K_a = \frac{1}{2}mu_b{}^2 - \frac{1}{2}mu_a{}^2$.

It is natural to try the same procedure starting with the relativistic expression for momentum $\mathbf{p} = m_0\mathbf{u}/\sqrt{1 - u^2/c^2}$.

$$K_b - K_a = \int_a^b \frac{d\mathbf{p}}{dt} \cdot d\mathbf{r}$$
$$= \int_a^b \frac{d}{dt}\left[\frac{m_0\mathbf{u}}{\sqrt{1 - u^2/c^2}}\right] \cdot \mathbf{u}\, dt$$
$$= \int_a^b \mathbf{u} \cdot d\left[\frac{m_0\mathbf{u}}{\sqrt{1 - u^2/c^2}}\right]$$

The integrand is $\mathbf{u} \cdot d\mathbf{p} = d(\mathbf{u} \cdot \mathbf{p}) - \mathbf{p} \cdot d\mathbf{u}$. Therefore

$$K_b - K_a = (\mathbf{u} \cdot \mathbf{p})\Big|_a^b - \int_a^b \mathbf{p} \cdot d\mathbf{u}$$
$$= \frac{m_0u^2}{\sqrt{1 - u^2/c^2}}\Bigg|_a^b - \int_a^b \frac{m_0u\,du}{\sqrt{1 - u^2/c^2}},$$

where we have used the earlier identity $\mathbf{u} \cdot d\mathbf{u} = u\,du$. The integral is elementary, and we find

$$K_b - K_a = \frac{m_0u^2}{\sqrt{1 - u^2/c^2}}\Bigg|_a^b + m_0c^2\sqrt{1 - \frac{u^2}{c^2}}\,\Bigg|_a^b.$$

Take point b as arbitrary, and let the particle be at rest at point a, $u_a = 0$.

$$K = \frac{m_0u^2}{\sqrt{1 - u^2/c^2}} + m_0c^2\sqrt{1 - \frac{u^2}{c^2}} - m_0c^2$$
$$= \frac{m_0[u^2 + c^2(1 - u^2/c^2)]}{\sqrt{1 - u^2/c^2}} - m_0c^2$$
$$= \frac{m_0c^2}{\sqrt{1 - u^2/c^2}} - m_0c^2$$

or

$$K = mc^2 - m_0c^2, \qquad 13.3$$

where $m = m_0/\sqrt{1 - u^2/c^2}$.

This expression for kinetic energy bears little resemblance to its classical counterpart. However, in the limit $u \ll c$, the relativistic result should approach the classical expression $K = \frac{1}{2}mu^2$. This is indeed the case, as we see by making the approximation $1/\sqrt{1 - u^2/c^2} \approx 1 + \frac{1}{2}u^2/c^2$. Then

$$\begin{aligned} K &= \frac{m_0c^2}{\sqrt{1 - u^2/c^2}} - m_0c^2 \\ &\approx m_0c^2\left(1 + \frac{1}{2}\frac{u^2}{c^2} - 1\right) \\ &= \tfrac{1}{2}m_0u^2. \end{aligned}$$

The kinetic energy arises from the work done on the particle to bring it from rest to speed u. Suppose that we rewrite Eq. (13.3) as

$$\begin{aligned} mc^2 &= K + m_0c^2 \\ &= \text{work done on particle} + m_0c^2. \end{aligned} \qquad 13.4$$

Einstein proposed the following bold interpretation of this result: mc^2 is the *total* energy E of the particle. The first term arises from external work; the second term, m_0c^2, represents the "rest" energy the particle possesses by virtue of its mass. In summary

$$E = mc^2. \qquad 13.5$$

It is important to realize that Einstein's generalization goes far beyond the classical conservation law for mechanical energy. Thus, if energy ΔE is added to a body, its mass will change by $\Delta m = \Delta E/c^2$, irrespective of the form of energy. ΔE could represent mechanical work, heat energy, the absorption of light, or any other form of energy. In relativity the classical distinction between mechanical energy and other forms of energy disappears. Relativity treats all forms of energy on an equal footing, in contrast to classical physics where each form of energy must be treated as a special case. The conservation of total energy $E = mc^2$ is a consequence of the very structure of relativity. In the next chapter we shall show that the conservation laws for energy and momentum are different aspects of a single, more general, conservation law.

The following example illustrates the relativistic concept of energy and the validity of the conservation laws in different inertial frames.

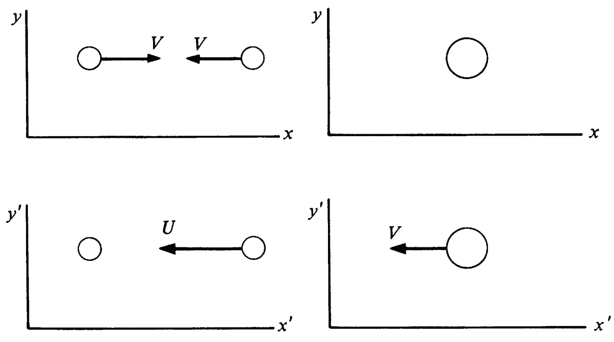
Example 13.2 **Relativistic Energy and Momentum in an Inelastic Collision**

Suppose that two identical particles collide with equal and opposite velocities and stick together. Classically, the initial kinetic energy is $2(\frac{1}{2}MV^2) = MV^2$, where M is the newtonian mass. By conservation of momentum the mass $2M$ is at rest and has zero kinetic energy. In the language of Chap. 4 we say that mechanical energy MV^2 was lost as heat. As we shall see, this distinction between forms of energy does not occur in relativity.

Now consider the same collision relativistically, as seen in the original frame x, y, and in a frame x', y' moving with one of the particles. By the relativistic transformation of velocities, Eq. (12.2),

$$U = \frac{2V}{1 + V^2/c^2} \qquad 1$$

in the direction shown.



Let the rest mass of each particle be M_{0i} before the collision and M_{0f} after the collision. In the x, y frame, momentum is obviously conserved. The total energy before the collision is $2M_{0i}c^2/\sqrt{1 - V^2/c^2}$, and after the collision the energy is $2M_{0f}c^2$. No external work was done on the particles, and the total energy is unchanged. Therefore,

$$\frac{2M_{0i}c^2}{\sqrt{1 - V^2/c^2}} = 2M_{0f}c^2$$

or

$$M_{0f} = \frac{M_{0i}}{\sqrt{1 - V^2/c^2}}. \qquad 2$$

The final rest mass is greater than the initial rest mass because the particles are warmer. To see this, we take the low velocity approximation

$$M_{0f} \approx M_{0i}\left(1 + \frac{1}{2}\frac{V^2}{c^2}\right)$$

The increase in rest energy for the two particles is $2(M_{0f} - M_{0i})c^2 \approx 2(\frac{1}{2}M_{0i}V^2)$, which corresponds to the loss of classical kinetic energy. Now, however, the kinetic energy is not "lost"—it is present as a mass increase.

By the postulate that all inertial frames are equivalent, the conservation laws must hold in the x', y' frame as well. If our assumed conservation laws possess this necessary property, we have in the x', y' frame

$$\frac{M_{0i}U}{\sqrt{1 - U^2/c^2}} = \frac{2M_{0f}V}{\sqrt{1 - V^2/c^2}} \qquad 3$$

by conservation of momentum and

$$M_{0i}c^2 + \frac{M_{0i}c^2}{\sqrt{1 - U^2/c^2}} = \frac{2M_{0f}c^2}{\sqrt{1 - V^2/c^2}} \qquad 4$$

by the conservation of energy.

The question now is whether Eqs. (3) and (4) are consistent with our earlier results, Eqs. (1) and (2). To check Eq. (3), we use Eq. (1) to write

$$\begin{aligned} 1 - \frac{U^2}{c^2} &= 1 - \frac{4V^2/c^2}{(1 + V^2/c^2)^2} \\ &= \frac{(1 - V^2/c^2)^2}{(1 + V^2/c^2)^2}. \end{aligned} \qquad 5$$

From Eqs. (1) and (5),

$$\begin{aligned} \frac{U}{\sqrt{1 - U^2/c^2}} &= \frac{2V}{(1 + V^2/c^2)}\frac{(1 + V^2/c^2)}{(1 - V^2/c^2)} \\ &= \frac{2V}{1 - V^2/c^2} \end{aligned}$$

and the left hand side of Eq. (3) becomes

$$\frac{M_{0i}U}{\sqrt{1 - U^2/c^2}} = \frac{2M_{0i}V}{1 - V^2/c^2}. \qquad 6$$

From Eq. (2), $M_{0i} = M_{0f}\sqrt{1 - V^2/c^2}$, and Eq. (6) reduces to

$$\frac{M_{0i}U}{\sqrt{1 - U^2/c^2}} = \frac{2M_{0f}V}{\sqrt{1 - V^2/c^2}},$$

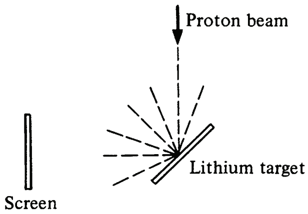
which is identical to Eq. (3). Similarly, it is not hard to show that Eq. (4) is also consistent.

We see from Eq. (6) that if we had assumed that rest mass was unchanged in the collision, $M_{0i} = M_{0f}$, the conservation law for momentum (or for energy) would not be correct in the second inertial frame. The relativistic description of energy plays an essential part in maintaining the validity of the conservation laws in all inertial frames.

Example 13.3 The Equivalence of Mass and Energy

In 1932 Cockcroft and Walton, two young British physicists, successfully operated the first high energy proton accelerator and succeeded in causing a nuclear disintegration. Their experiment provided one of the earliest confirmations of the relativistic mass-energy relation.

Briefly, their accelerator consisted of a power supply that could reach 600 kV and a source of protons (hydrogen nuclei). The power supply used an ingenious arrangement of capacitors and rectifiers to quadruple the voltage of a 150-kV supply. The protons were supplied by an electrical discharge in hydrogen and were accelerated in vacuum by the applied high voltage.



Cockcroft and Walton studied the effect of the protons on a target of ^{7}Li (lithium, having atomic mass 7). A zinc sulfide fluorescent screen, located nearby, emitted occasional flashes, or scintillations. By various tests they determined that the scintillations were due to alpha particles, the nuclei of helium, ^{4}He. Their interpretation was that the ^{7}Li captures a proton and that the resulting nucleus of mass 8 immediately disintegrates into two alpha particles. We can write the reaction as

$$^1\text{H} + {}^7\text{Li} \rightarrow {}^4\text{He} + {}^4\text{He}.$$

The mass energy equation for the reaction is

$$K(^1\text{H}) + [M(^1\text{H}) + M(^7\text{Li})]c^2 = 2K(^4\text{He}) + 2M(^4\text{He})c^2$$

where $K(^1\text{H})$ is the kinetic energy of the incident proton, $K(^4\text{He})$ is the kinetic energy of each of the emitted alpha particles, and $M(^1\text{H})$ is the proton rest mass, etc. (The initial momentum of the proton is negligible, and the two alpha particles are emitted back to back with equal energy by conservation of momentum.)

We can rewrite the mass-energy equation as

$$K = \Delta M c^2,$$

where $K = 2K(^4\text{He}) - K(^1\text{H})$, and ΔM is the initial rest mass minus the final rest mass.

The energy of the alpha particles was determined by measuring their range. Cockcroft and Walton obtained the value $K = 17.2$ MeV (1 MeV $= 10^6$ eV $= 1.6 \times 10^{-13}$ J).

The relative masses of the nuclei were known from mass spectrometer measurements. In atomic mass units, amu, defined so that $M(^{16}\text{O}) = 16$, the values available to Cockcroft and Walton were

$$M(^1\text{H}) = 1.0072$$
$$M(^7\text{Li}) = 7.0104 \pm 0.0030$$
$$M(^4\text{He}) = 4.0011.$$

These yield

$$\Delta M = (1.0072 + 7.0104) - 2(4.0011)$$
$$= (0.0154 \pm 0.0030) \text{ amu.}$$

The rest energy of 1 amu is 931 MeV and therefore

$$\Delta M c^2 = (14.3 \pm 2.7) \text{ MeV.}$$

The difference between K and $\Delta M c^2$ is (17.2 − 14.3) MeV = 2.9 MeV, slightly larger than the experimental uncertainty of 2.7 MeV. However, the experimental uncertainty always represents an estimate, not a precise limit, and the result can be taken as consistent with the relation $K = \Delta M c^2$.

It is clear that the masses must be known to high accuracy for studying the energy balance in nuclear reactions. Modern techniques of mass spectrometry have achieved an accuracy of better than 10^{-5} amu, and the mass-energy equivalence has been amply confirmed. According to a modern table of masses, the decrease in rest mass in the reaction studied by Cockcroft and Walton is $\Delta M c^2 = (17.3468 \pm 0.0012)$ MeV.

Often it is useful to express the total energy of a free particle in terms of its momentum. Classically the relation is

$$E = \tfrac{1}{2}mv^2 = \frac{p^2}{2m}.$$

To find the equivalent relativistic expression we must combine the relativistic momentum

$$\mathbf{p} = m\mathbf{u} = \frac{m_0\mathbf{u}}{\sqrt{1 - u^2/c^2}} = m_0\mathbf{u}\gamma \qquad 13.6$$

with the energy

$$E = mc^2 = m_0c^2\gamma. \qquad 13.7$$

Squaring Eq. (13.6) gives

$$p^2 = \frac{m_0{}^2u^2}{1 - u^2/c^2},$$

which we solve for γ as follows:

$$\frac{u^2}{c^2} = \frac{p^2}{p^2 + m_0{}^2c^2}$$

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

$$= \sqrt{1 + \frac{p^2}{m_0{}^2c^2}}.$$

Inserting this in Eq. (13.7), we have

$$E = m_0c^2\sqrt{1 + \frac{p^2}{m_0{}^2c^2}}.$$

The square of this equation is algebraically somewhat simpler and is the form usually employed.

$$E^2 = (pc)^2 + (m_0c^2)^2 \qquad 13.8$$

We have derived the relativistic expressions for momentum and energy by invoking conservation laws. However, we have not dealt with the role of force in relativity. It is possible to attack this problem by considering the form of the equations of motion in various coordinate systems. We shall develop a systematic way of doing this in the next chapter, and so we defer the problem of force for the present.

For convenience, here is a summary of the important dynamical formulas we have developed so far.

$$\mathbf{p} = m\mathbf{u} = m_0\mathbf{u}\gamma \qquad 13.9$$

$$K = mc^2 - m_0c^2 = m_0c^2(\gamma - 1) \qquad 13.10$$

$$E = mc^2 = m_0c^2\gamma \qquad 13.11$$

$$E^2 = (pc)^2 + (m_0c^2)^2 \qquad 13.12$$

13.3 Massless Particles

A surprising consequence of the relativistic energy-momentum relation is the possibility of "massless" particles—particles which possess momentum and energy but no rest mass. If we take $m_0 = 0$ in the relation

$$E^2 = (pc)^2 + (m_0c^2)^2,$$

the result is

$$E = pc. \tag{13.13}$$

We take the positive root on the plausible assumption that particles whose energy decreases with increasing momentum would be unstable.

In order to have nonzero momentum we must have a finite value for

$$\mathbf{p} = m_0\mathbf{u}\Big/\sqrt{1 - \frac{u^2}{c^2}}.$$

in the limit $m_0 \to 0$. This is only possible if $u \to c$ as $m_0 \to 0$; massless particles must travel at the speed of light.

The principal massless particle known to physics is the *photon*, the particle of light. Photons interact electromagnetically with electrons and other charged particles and are easy to detect with photographic films, phototubes, or the eye. The *neutrino*, which is associated with the weak forces of radioactive beta decay, is believed to be massless, but it interacts so weakly with matter that its direct detection is extremely difficult. (The sun is a copious source of neutrinos, but most of the solar neutrinos which reach the earth pass through it without interacting.) Experiments have shown that the neutrino rest mass is no larger than 1/2,000 the rest mass of the electron, and it could well be zero. There are theoretical reasons for believing in the existence of the graviton, a massless particle associated with the gravitational force. The graviton's interaction with matter is so weak that it has not yet been detected.

We owe the concept of the photon to Einstein, who introduced it in his pioneering paper on the photoelectric effect published a few months before his work on relativity.[1] Briefly, Einstein proposed that the energy of a light wave can only be transmitted to matter in discrete amounts, or quanta, of value $h\nu$, where h is Planck's constant 6.63×10^{-34} J/Hz, and ν is the frequency of the light wave in hertz. The arguments for this proposal grew out of Einstein's concern with problems in classical electromagnetic theory and considerations of Planck's quantum hypothesis,

[1] Within a period of one year Einstein wrote four papers, each of which became a classic, on the photoelectric effect, relativity, brownian motion, and the quantum theory of the heat capacity of solids. It was for his work on the photoelectric effect, not relativity, that Einstein received the Nobel Prize for Physics in 1921. Relativity was so encumbered with philosophical and political implications that the Nobel committee refused to acknowledge it. This regrettable incident was unique in the history of the prize.

a theory constructed by Planck in 1900 to overcome difficulties in classical statistical mechanics. Although we cannot develop here the background necessary to justify Einstein's theory of the photon, perhaps the following experimental evidence will help make the photon seem plausible.

Example 13.4 The Photoelectric Effect

In 1887 Heinrich Hertz discovered that metals can give off electrons when illuminated by ultraviolet light. This process, the *photoelectric effect,* represents the direct conversion of light into mechanical energy (here, the kinetic energy of the electron). Einstein predicted that the energy an electron absorbs from a beam of light at frequency ν is exactly $h\nu$. If the electron loses a certain amount of energy W in leaving the metal, then the kinetic energy of the emitted electron is

$$K = h\nu - W.$$

W is known as the *work function* of the metal. The work function is typically a few electron volts, but unfortunately it depends on the chemical state of the metal surface, making the photoelectric effect a difficult matter to investigate. Millikan overcame this problem in 1916 by working with metal surfaces prepared in a high vacuum system. The kinetic energy was determined by measuring the photocurrent collected on a plate near the metal and applying an electric potential between the plate and photosurface just adequate to stop the current. If the potential is $-V$, then the energy lost by the electrons as they travel to the plate is $(-e)(-V)$. At cutoff we have $V = V_c$ and

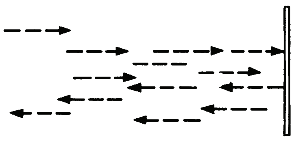
$$eV_c = h\nu - W.$$

Millikan observed the cutoff voltage as a function of frequency for several alkali metals. In accord with Einstein's formula, he found that V_c was a linear function of ν, with slope h/e, and that V_c was independent of the intensity of the light.

If the energy of light were absorbed by the electron according to the classical picture, the electrons would have a wide energy distribution depending on the intensity of the light, in sharp disagreement with Millikan's results. The fact that light can interfere with itself, as in the Michelson interferometer, is compelling evidence that light has wave properties. Nevertheless, the photoelectric effect illustrates that light also has particle properties. Einstein's energy relation, $E = h\nu$, provides the link between these apparently conflicting descriptions of light by relating the energy of the particle to the frequency of the wave.

Example 13.5 Radiation Pressure of Light

A consequence of Maxwell's electromagnetic theory is that a light wave carries momentum which it will transfer to a surface when it is reflected or absorbed. The result, as we know from our study of momentum in Chap. 3, is a pressure on the surface. The calculation of radiation pressure is complicated using the wave theory of light, but with the photon picture it is simple.



Consider a stream of photons striking a perfectly reflecting mirror at normal incidence. The initial momentum of each photon is $p = E/c$, and the total change in momentum in the reflection is $2p = 2E/c$. If there are n photons incident per unit area per second, the total momentum change per second is $2nE/c$, and this is equal to the force per unit area exerted on the mirror by the light. Hence the radiation pressure P is

$$P = \frac{2nE}{c} = \frac{2I}{c},$$

where $I = nE$ is the intensity of the light, the power per unit area. Similarly, the radiation pressure on a perfect absorber is I/c.

The average intensity of sunlight falling on the earth's surface at normal incidence, known as the *solar constant,* is $I \approx 1{,}000$ W/m^2. The radiation pressure on a mirror due to sunlight is therefore $P = 2I/c = 7 \times 10^{-6}$ N/m^2, a very small pressure. (Atmospheric pressure is 10^5 N/m^2.) On the cosmic scale, however, radiation pressure is large; it helps keep stars from collapsing under their own gravitational forces.

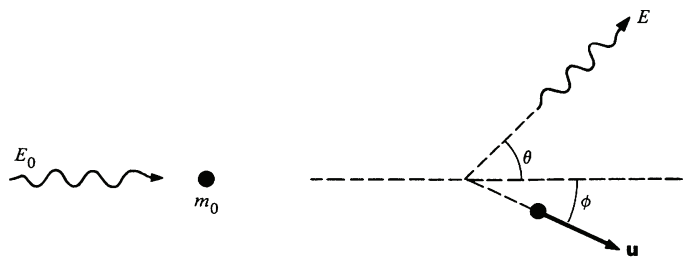
Since the photon is a completely relativistic particle, newtonian physics provides little insight into its properties. For instance, unlike classical particles, photons can be created and destroyed; the absorption of light by matter corresponds to the destruction of photons, while the process of radiation corresponds to the creation of photons. Nevertheless, the familiar laws of conservation of momentum and energy, as generalized in the theory of relativity, are sufficiently powerful to let us draw conclusions about processes involving photons without a detailed knowledge of the interaction, as the following examples illustrate.

Example 13.6 The Compton Effect

The special theory of relativity was not widely accepted by the 1920s partly because of the radical nature of its concepts, but also because there was little experimental evidence. In 1922 Arthur Compton performed a refined experiment on the scattering of x-rays from matter which left little doubt that relativistic dynamics was valid.

A photon of visible light has energy in the range of 1 to 2 eV, but photons of much higher energy can be obtained from x-ray machines, radioactive sources, or particle accelerators. X-ray photons have energies typically in the range 10 to 100 keV, and their wavelengths can be measured with high accuracy by the technique of crystal diffraction.

When a photon collides with a free electron, the conservation laws require that the photon lose a portion of its energy. The outgoing photon therefore has a longer wavelength than the primary photon, and this shift in wavelength, first observed by Compton, is known as the *Compton effect.*



Let the photon have initial energy E_0 and momentum E_0/c, and suppose that the electron is initially at rest. After the collision, the electron is scattered at angle ϕ with velocity $\mathbf{u}$ and the photon is scattered at angle θ with energy E. Let $E_e = m_0c^2/\sqrt{1 - u^2/c^2}$ be the final electron energy and $\mathbf{p} = m\mathbf{u}$ the momentum. Then, by conservation of energy,

$$E_0 + m_0c^2 = E + E_e. \qquad 1$$

By conservation of momentum,

$$\frac{E_0}{c} = \frac{E}{c}\cos\theta + p\cos\phi \qquad 2$$

$$0 = \frac{E}{c}\sin\theta - p\sin\phi. \qquad 3$$

Our object is to eliminate reference to the electron and find E as a function of θ, since Compton detected only the outgoing photon in his experiments. Equations (2) and (3) can be written

$$(E_0 - E\cos\theta)^2 = (pc)^2\cos^2\phi$$
$$(E\sin\theta)^2 = (pc)^2\sin^2\phi.$$

Adding,

$$E_0{}^2 - 2E_0E\cos\theta + E^2 = (pc)^2 \qquad 4$$
$$= E_e{}^2 - (m_0c^2)^2,$$

where we have used the energy-momentum relation, Eq. (13.12). Using Eq. (1) to eliminate E_e, Eq. (4) becomes

$$E_0{}^2 - 2E_0E\cos\theta + E^2 = (E_0 + m_0c^2 - E)^2 - (m_0c^2)^2,$$

which reduces to

$$E = \frac{E_0}{1 + (E_0/m_0c^2)(1 - \cos\theta)}. \qquad 5$$

Note that E is always greater than zero, which means that a free electron cannot absorb a photon.

Compton measured wavelengths rather than energies in his experiment. From the Einstein frequency condition, $E_0 = h\nu_0 = hc/\lambda_0$ and $E = hc/\lambda$, where λ_0 and λ are the wavelengths of the incoming and outgoing photons, respectively. In terms of wavelength, Eq. (5) takes the simple form

$$\lambda = \lambda_0 + \frac{h}{m_0c}(1 - \cos\theta).$$

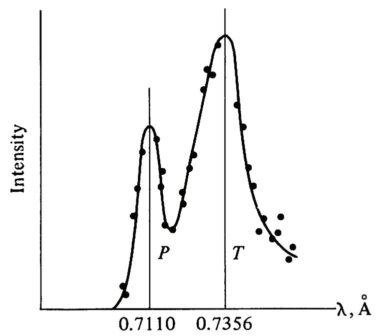
The quantity h/m_0c is known as the *Compton wavelength* of the electron and has the value

$$\frac{h}{m_0c} = 2.426 \times 10^{-12}\text{ m}$$
$$= 0.02426\text{ Å},$$

where 1 Å $= 10^{-10}$ m.

The shift in wavelength at a given angle is independent of the initial photon energy:

$$\lambda - \lambda_0 = \frac{h}{m_0c}(1 - \cos\theta).$$



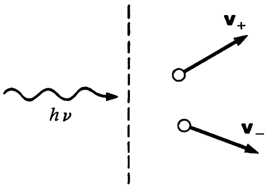
The figure shows one of Compton's results for $\lambda_0 = 0.711$ Å and $\theta = 90°$, where peak P is due to primary photons and peak T to the Compton scattered photons from a block of graphite. The measured wavelength shift is approximately 0.0246 Å and the calculated value is 0.02426 Å. The difference is less than the estimated uncertainty due to the limited resolution of the spectrometer and other experimental limitations.

In our analysis we assumed that the electron was free and at rest. For sufficiently high proton energies, this is a good approximation for electrons in the outer shells of light atoms. If the motion of the electrons is taken into account, the Compton peak is broadened. (The broadening of peak T in the figure compared with P shows this effect.)

If the binding energy of the electron is comparable to the photon energy, momentum and energy can be transferred to the atom as a whole, and the photon can be completely absorbed.

Example 13.7 Pair Production

We have already seen two ways by which a photon can lose energy in matter, photoelectric absorption and Compton scattering. If a photon's energy is sufficiently high, it can also lose energy in matter by the mechanism of *pair production*. The rest mass of an electron is $m_0c^2 =$ 0.511 MeV. Can a photon of this energy create an electron? The answer is no, since this would require the creation of a single electric charge. As far as we know, electric charge is conserved in all physical processes. However, if equal amounts of positive and negative charge are created, the total charge remains zero and charge is conserved. Hence, it is possible to create an electron-positron pair (e^--e^+), two particles having the same mass but opposite charge.



A single photon of energy $2m_0c^2$ or greater has enough energy to form an e^--e^+ pair, but the process cannot occur in free space because it would not conserve momentum. If we imagine that the process occurs, conservation of energy gives

$$h\nu = m_+c^2 + m_-c^2 = (\gamma_+ + \gamma_-)m_0c^2,$$

or

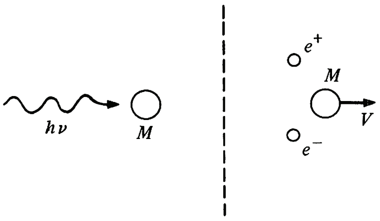
$$\frac{h\nu}{c} = (\gamma_+ + \gamma_-)m_0c,$$

while conservation of momentum gives

$$h\nu/c = |\gamma_+\mathbf{v}_+ + \gamma_-\mathbf{v}_-|m_0.$$

These equations cannot be satisfied simultaneously because

$$(\gamma_+ + \gamma_-)c > |\gamma_+\mathbf{v}_+ + \gamma_-\mathbf{v}_-|.$$



Pair production is possible if a third particle is available for carrying off the excess momentum. For instance, suppose that the photon hits a nucleus of rest mass M and creates an e^--e^+ pair at rest. We have

$$h\nu + Mc^2 = 2m_0c^2 + Mc^2\gamma.$$

Since nuclei are much more massive than electrons, let us assume that $h\nu \ll Mc^2$. (For hydrogen, the lightest atom, this means that $h\nu \ll$ 940 MeV.) In this case the atom will not attain relativistic speeds and we can make the classical approximation

$$\begin{aligned} h\nu &= 2m_0c^2 + Mc^2(\gamma - 1) \\ &\approx 2m_0c^2 + \tfrac{1}{2}MV^2. \end{aligned}$$

To the same approximation, conservation of momentum yields

$$\frac{h\nu}{c} = MV.$$

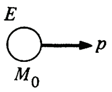
Substituting this in the energy expression gives

$$h\nu = 2m_0c^2 + \frac{1}{2}\frac{(h\nu)^2}{Mc^2} \approx 2m_0c^2,$$

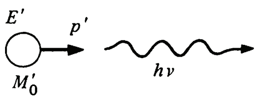
since we have already assumed $h\nu \ll Mc^2$. The threshold for pair production in matter is therefore $2m_0c^2 = 1.02$ MeV. The nucleus plays an essentially passive role, but by providing for momentum conservation it allows an otherwise forbidden process to occur.

Example 13.8 **The Photon Picture of the Doppler Effect**

In Chap. 12 we discussed the Doppler effect from the standpoint of wave theory, but we can also treat it using the photon picture. Consider first an atom with rest mass M_0, held stationary. If the atom emits a photon of energy $h\nu_0$, its new rest mass is given by $M_0'c^2 = M_0c^2 - h\nu_0$.



Suppose now that the atom moves freely with velocity **u** before emitting the photon. The atom's energy is $E = M_0c^2/\sqrt{1 - u^2/c^2}$ and its momentum is $p = M_0u/\sqrt{1 - u^2/c^2}$. After the emission of a photon of energy $h\nu$ the atom has velocity **u**$'$, rest mass M_0', energy E', and momentum p'. For simplicity, we consider the photon to be emitted along the line of motion. By conservation of energy and momentum we have



$$E = E' + h\nu \qquad 1$$

$$p = p' + \frac{h\nu}{c}. \qquad 2$$

Therefore,

$$(E - h\nu)^2 = E'^2$$

$$(pc - h\nu)^2 = (p'c)^2$$

and

$$(E - h\nu)^2 - (pc - h\nu)^2 = E'^2 - (p'c)^2 = (M_0'c^2)^2 \qquad 3$$

by the energy-momentum relation. Expanding the left hand side and using $E^2 - (pc)^2 = (M_0c^2)^2$, we obtain

$$\begin{aligned}(M_0c^2)^2 - 2Eh\nu + 2pch\nu &= (M_0'c^2)^2\\ &= (M_0c^2 - h\nu_0)^2.\end{aligned}$$

Simplifying, we find that

$$\nu = \nu_0 \frac{(2M_0c^2 - h\nu_0)}{2(E - pc)}.$$

However,

$$\begin{aligned}E - pc &= \frac{M_0c^2}{\sqrt{1 - u^2/c^2}}\left(1 - \frac{u}{c}\right)\\ &= M_0c^2\sqrt{\frac{1 - u/c}{1 + u/c}}.\end{aligned}$$

Hence,

$$\nu = \nu_0 \left(1 - \frac{h\nu_0}{2M_0c^2}\right) \sqrt{\frac{1 + u/c}{1 - u/c}}.$$

The term $h\nu_0/2M_0c^2$ represents a decrease in the photon energy due to the recoil energy of the atom. For a massive source, this term is negligible and

$$\nu = \nu_0 \sqrt{\frac{1 + u/c}{1 - u/c}},$$

in agreement with the result of the last chapter, Eq. (12.6).

Although it is always satisfying to derive a result by different arguments, perhaps the chief interest in this exercise is to show how two completely different views of light, wave and particle, lead to exactly the same prediction for the shift in frequency of radiation from a moving source.

13.4 Does Light Travel at the Velocity of Light?

Although the title of this section may sound rhetorical, the question is not trivial. It is apparent that the velocity of light plays a special role in relativity. In fact, Einstein created the special theory of relativity primarily from considerations of Maxwell's electromagnetic theory, the theory of light. However, it is important to realize that the real significance of the velocity of light is that it exemplifies a *universal* velocity, a velocity whose value is the same for an observer in any inertial system. There can be only *one* such universal velocity in the theory of relativity, as the following argument shows.

Suppose that there is a second universal velocity c^* representing the velocity of some phenomenon other than light—perhaps the speed of gravitons or neutrinos. Let us call the phenomenon Γ. Consider a light pulse and a Γ pulse emitted along the x axis from the origin of the x, y system at $t = 0$. The pulses travel according to:

Light: $x_l = ct$

Γ: $x_\Gamma = c^*t.$

The relative velocity of the two pulses is

$$u = \frac{d}{dt}(x_\Gamma - x_l)$$
$$= c^* - c.$$

Now consider the same pulses in the x', y' system which is moving along the x axis with velocity V. Since c^* and c are universal velocities, the loci of the pulses must be given by

$$x_l' = ct'$$
$$x_\Gamma' = c^*t'.$$

The relative velocities of the two pulses is

$$u' = \frac{d}{dt'}[x_\Gamma' - x_l']$$
$$= c^* - c,$$

as before. But the relativistic transformation of velocities gives

$$c' = \frac{c - V}{1 - cV/c^2} = c$$
$$(c^*)' = \frac{c^* - V}{1 - c^*V/c^2}.$$

Thus, the Lorentz transformations predict that

$$u' = (c^*)' - c$$
$$= \frac{c^* - V}{1 - c^*V/c^2} - c.$$

This disagrees with the result above, $u' = c^* - c$, unless $c^* = c$, in which case $u = 0$ and $u' = 0$. We conclude that there can be only one universal velocity.

If this argument seems rather formal, perhaps the following explanation will help. The theory of relativity satisfies the postulate of relativity: all inertial coordinate systems are equivalent. It also satisfies the postulate that the velocity of light is a universal constant: all observers in inertial systems will obtain the same result for the velocity of a particular light signal. However, the theory of relativity cannot accommodate more than one such universal velocity; if we try to introduce a second universal velocity, the whole edifice of relativity collapses. In particular, we can no longer obtain a consistent recipe for relating coordinates of events in different systems.

With this background, perhaps we can rephrase the title of this section more meaningfully as "does light travel with the universal velocity?" The question is actually quite interesting and a matter of current investigation.

Example 13.9 **The Rest Mass of the Photon**

If the photon had a nonzero rest mass, the velocity of light would differ from c. If we let m_p represent the rest mass of a photon, we would have

$$E = \gamma m_p c^2.$$

If we assume that the photon energy-frequency relation $E = h\nu$ remains valid, then squaring the equation above gives

$$(h\nu)^2 = (m_p c^2)^2 \frac{1}{1 - v^2/c^2},$$

or, after rearranging,

$$\frac{v^2}{c^2} = 1 - \frac{\nu_0{}^2}{\nu^2},$$

where $h\nu_0 = m_p c^2$. ν_0 plays the role of a characteristic frequency of the photon: $h\nu_0$ is the rest energy of the photon. If $\nu_0 = 0$, we have $v = c$. Otherwise, the velocity of light depends on frequency. Behavior such as this is well known when light passes through a refractive medium such as glass or water; it is known as *dispersion.* The question for experiment to decide is whether or not empty space exhibits dispersion.

There have been a number of recent attempts to set a limit on the rest mass of the photon (or, better still, to measure it, although at present there is no compelling reason to believe that the rest mass is not zero).

Example 13.10 **Light from a Pulsar**

Pulsars are stars that emit regular bursts of energy at repetition frequencies from 30 to 0.1 Hz. They were discovered in 1968 and their unexpected properties have been a source of much excitement among astronomers and astrophysicists. Perhaps the most interesting pulsar is the one in the Crab nebula. It has the highest frequency, 30 Hz, and is the only one so far observed which pulses in the optical and x-ray regions, as well as at radio frequencies. The pulses are quite sharp, and their arrival time can be measured to an accuracy of microseconds. It is known that light from the pulsar at different optical wavelengths arrives simultaneously within the experimental resolving time. We can use these facts to set a limit on the rest mass of the photon.

It takes light 5,000 years to reach us from the Crab nebula. Suppose that signals at two different frequencies travel with a small difference in velocity, Δv, and arrive at slightly different times, T and $T + \Delta T$. Since $T = L/v$, where L is the distance from the Crab nebula, we have

$$\Delta v = -\frac{L}{T^2}\Delta T$$

or

$$\frac{\Delta v}{v} = -\frac{\Delta T}{T}.$$

No such velocity difference has been observed, but by estimating the sensitivity of the experiment we can set an upper limit to Δv. ΔT can be measured to an accuracy of about 2×10^{-3} s, and using $T = 5 \times 10^3$ years $= 1.5 \times 10^{11}$ s, we have

$$\left|\frac{\Delta v}{c}\right| = \left|\frac{\Delta T}{T}\right| < \frac{2 \times 10^{-3}}{1.5 \times 10^{11}}$$
$$\approx 10^{-14},$$

where we have taken $v \approx c$.

Now let us translate this limit on Δv into a limit on the possible rest mass of a photon. From the result of the last example,

$$\frac{v^2}{c^2} = 1 - \frac{\nu_0{}^2}{\nu^2}.$$

Consider signals at two different frequencies, ν_1 and ν_2. We have

$$\frac{v_1{}^2 - v_2{}^2}{c^2} = \nu_0{}^2\left(\frac{1}{\nu_2{}^2} - \frac{1}{\nu_1{}^2}\right).$$

The left hand side can be written

$$\frac{(v_1 - v_2)(v_1 + v_2)}{c^2} \approx 2\frac{\Delta v}{c},$$

where we have taken $(v_1 - v_2) = \Delta v$, and $v_1 + v_2 \approx 2c$. For observations made in the optical region we can take $\nu_1 = 8 \times 10^{14}$ Hz (blue) and $\nu_2 = 5 \times 10^{14}$ Hz (red). Then, using the limit $\Delta v/c < 2 \times 10^{-16}$, we have

$$2 \times 2 \times 10^{-16} > \frac{\nu_0{}^2}{10^{28}}\left(\frac{1}{5^2} - \frac{1}{8^2}\right) = 2.4 \times 10^{-30}\nu_0{}^2$$

or

$$\nu_0 < 10^7 \text{ Hz}.$$

This gives an upper limit to the photon rest mass of

$$m_p = \frac{h\nu_0}{c^2} < 10^{-40} \text{ kg}.$$

An even lower limit to the photon rest mass can be found by observing the arrival time of radio pulses from the Crab nebula. The analysis is somewhat more complicated because of the effect of free electrons in interstellar space. The result is that the rest mass of the photon has an upper limit of 10^{-47} kg.

Problems

13.1 It is estimated that a cosmic ray primary proton can have energy up to 10^{13} MeV (almost 10^8 greater than the highest energy achieved with an accelerator). Our galaxy has a diameter of about 10^5 light-years. How long does it take the proton to traverse the galaxy, in its own rest frame? (1 eV $= 1.6 \times 10^{-19}$ J, $M_p = 1.67 \times 10^{-27}$ kg.)

13.2 When working with particles it is important to know when relativistic effects have to be considered.

A particle of rest mass m_0 is moving with speed v. Its classical kinetic energy is $K_{cl} = m_0v^2/2$. Let K_{rel} be the relativistic expression for its kinetic energy.

a. By expanding K_{rel}/K_{cl} in powers of v^2/c^2, estimate the value of v^2/c^2 for which K_{rel} differs from K_{cl} by 10 percent.

b. For this value of v^2/c^2, what is the kinetic energy in MeV of
(1) An electron ($m_0c^2 = 0.51$ MeV)
(2) A proton ($m_0c^2 = 930$ MeV)

13.3 In newtonian mechanics, the kinetic energy of a mass m moving with velocity $\mathbf{v}$ is $K = mv^2/2 = p^2/(2m)$ where $\mathbf{p} = m\mathbf{v}$. Hence, the change in kinetic energy due to a small change in momentum is $dK = \mathbf{p} \cdot d\mathbf{p}/m = \mathbf{v} \cdot d\mathbf{p}$.

Show that the relation $dK = \mathbf{v} \cdot d\mathbf{p}$ also holds in relativistic mechanics.

13.4 Two particles of rest mass m_0 approach each other with equal and opposite velocity v, in the laboratory frame. What is the total energy of one particle as measured in the rest frame of the other?

Ans. clue. If $v^2/c^2 = \frac{1}{2}$, $E = 3m_0c^2$

13.5 A particle of rest mass m and speed v collides and sticks to a stationary particle of mass M. What is the final speed of the composite particle?

Ans. $v_f = \gamma vm/(\gamma m + M)$, where $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$

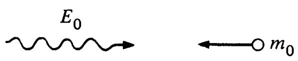
13.6 A particle of rest mass m_0 and kinetic energy xm_0c^2, where x is some number, strikes and sticks to an identical particle at rest. What is the rest mass of the resultant particle?

Ans. clue. If $x = 6$, $m = 4m_0$

13.7 In the laboratory frame a particle of rest mass m_0 and speed v is moving toward a particle of mass m_0 at rest.

What is the speed of the inertial frame in which the total momentum of the system is zero?

Ans. clue. If $v^2/c^2 = \frac{3}{4}$, the speed is $2v/3$



13.8 A photon of energy E_0 and wavelength λ_0 collides head on with a free electron of rest mass m_0 and speed v, as shown. The photon is scattered at 90°.

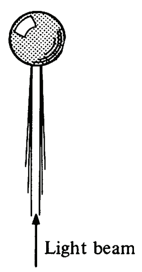
a. Find the energy E of the scattered photon.

Ans. $E = [E_0(1 + v/c)]/(1 + E_0/E_i)$, where $E_i = m_0c^2/\sqrt{1 - v^2/c^2}$

b. The outer electrons in a carbon atom move with speed $v/c \approx 6 \times 10^{-3}$. Using the result of part *a*, estimate the broadening in wavelength of the Compton scattered peak from graphite for $\lambda_0 = 0.711 \times 10^{-10}$ m and 90° scattering. The rest mass of an electron is 0.51 MeV and $h/(m_0c) = 2.426 \times 10^{-12}$ m. Neglect the binding of the electrons. Compare your result with Compton's data shown in Example 13.6.

13.9 The solar constant, the average energy per unit area falling on the earth, is 1.4×10^3 W/m². How does the total force of sunlight compare with the sun's gravitational force on the earth?

Sufficiently small particles can be ejected from the solar system by the radiation pressure of sunlight. Assuming a specific gravity of 5, what is the radius of the largest particle which can be ejected?



13.10 A 1-kW light beam from a laser is used to levitate a solid aluminum sphere by focusing it on the sphere from below. What is the diameter of the sphere, assuming that it floats freely in the light beam? The density of aluminum is 2.7 g/cm³.

13.11 A photon of energy E_0 collides with a free particle of mass m_0 at rest. If the scattered photon flies off at angle θ, what is the scattering angle of the particle, ϕ?

Ans. $\cot \phi = (1 + E_0/m_0c^2) \tan (\theta/2)$

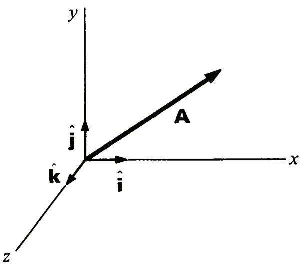
14 FOUR-VECTORS AND RELATIVISTIC INVARIANCE

14.1 Introduction

When a major advance in physics is made, old concepts inevitably lose importance and points of view which previously were of minor interest move to the center. Thus, with the advent of relativity the concept of the ether vanished, taking with it the problem of absolute motion. At the same time, the transformation properties of physical laws, previously of little interest, took on central importance. As we shall see in this chapter, transformation theory provides a powerful tool for generalizing nonrelativistic concepts and for testing the relativistic correctness of physical laws. Furthermore, it is a useful guide in the search for new laws. By using transformation theory we shall derive in a natural way the important results of relativity that we found by ad hoc arguments in the preceding chapters. This approach emphasizes the mathematical structure of physics and the nature of symmetry; it illustrates a characteristic mode of thought in contemporary physics.

To introduce the methods of transformation theory, we defer relativity for the moment and turn first to the transformation properties of ordinary vectors in three dimensions.

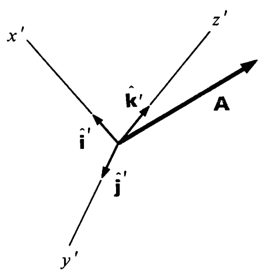
14.2 Vectors and Transformations



In Chap. 1 we defined vectors as "directed line segments"; with the help of transformation theory we can develop a more fundamental definition.

To motivate the argument and to illustrate the ideas of transformation theory we shall rely at first on our intuitive concept of vectors. Consider vector **A**, which represents some physical quantity such as force or velocity. To describe **A** in component form we introduce an orthogonal coordinate system x, y, z with unit base vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$. **A** can then be written

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}.$$



The coordinate system is not an essential part of the physics; it is a construct we introduce for convenience. We are perfectly free to use some other orthogonal coordinate system x', y', z' with base vectors $\hat{\mathbf{i}}'$, $\hat{\mathbf{j}}'$, $\hat{\mathbf{k}}'$. Let the x', y', z' system have the same origin as the x, y, z system, in which case the two systems are related by a rotation. In the primed system,

$$\mathbf{A} = A'_x\hat{\mathbf{i}}' + A'_y\hat{\mathbf{j}}' + A'_z\hat{\mathbf{k}}'.$$

For a general coordinate rotation, the components A'_x, A'_y, A'_z have a definite relation to the components A_x, A_y, A_z. Equating the two expressions for **A** gives

$$A'_x\hat{\mathbf{i}}' + A'_y\hat{\mathbf{j}}' + A'_z\hat{\mathbf{k}}' = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}.$$

If we take the dot product of both sides with $\hat{\mathbf{i}}'$ we obtain

$$A'_x = A_x(\hat{\mathbf{i}}' \cdot \hat{\mathbf{i}}) + A_y(\hat{\mathbf{i}}' \cdot \hat{\mathbf{j}}) + A_z(\hat{\mathbf{i}}' \cdot \hat{\mathbf{k}}) \qquad 14.1a$$

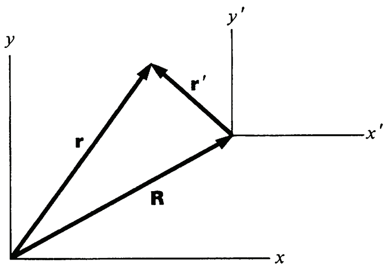
Similarly,

$$A'_y = A_x(\hat{\mathbf{j}}' \cdot \hat{\mathbf{i}}) + A_y(\hat{\mathbf{j}}' \cdot \hat{\mathbf{j}}) + A_z(\hat{\mathbf{j}}' \cdot \hat{\mathbf{k}}) \qquad 14.1b$$

$$A'_z = A_x(\hat{\mathbf{k}}' \cdot \hat{\mathbf{i}}) + A_y(\hat{\mathbf{k}}' \cdot \hat{\mathbf{j}}) + A_z(\hat{\mathbf{k}}' \cdot \hat{\mathbf{k}}). \qquad 14.1c$$

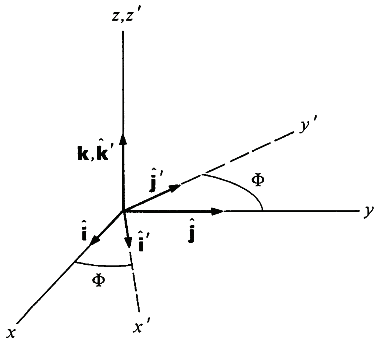
The coefficients $(\hat{\mathbf{i}}' \cdot \hat{\mathbf{i}})$, $(\hat{\mathbf{i}}' \cdot \hat{\mathbf{j}})$, etc., are numbers which are determined by the given rotation; they do not depend on **A**.

We derived Eq. (14.1) from our concept of vectors as directed line segments, but now we shall reverse the order and use Eq. (14.1) to define vectors. A vector in three dimensions is a set of three numbers which transform under a rotation of the coordinate system according to Eq. (14.1). It is easy to show that the vector algebra developed in Chap. 1 is consistent with our new definition of a vector. For example, the sum of two vectors is a vector, and the time derivative of a vector is also a vector.



$\mathbf{r}' = \mathbf{r} - \mathbf{R}$

We should point out that the general displacement of a coordinate system is composed of a translation as well as a rotation. The reason that we referred only to rotations in the definition of a vector is that translations have no effect on the components of a vector. The sole exception is the position vector **r**, which is defined with respect to a specific origin. The components of **r** transform under rotations according to Eq. (14.1), but **r** can be distinguished from true vectors such as **F** and **v** by its transformation properties under translation. We can distinguish between true vectors, position vectors, and other mathematical entities by investigating how they behave under all possible transformations.



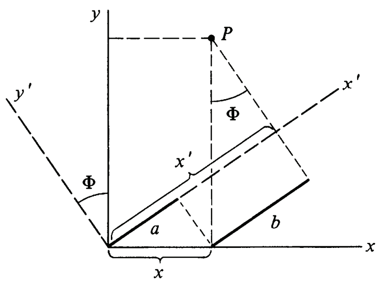
Rotation about the z axis

Equation (14.1) is completely general, but usually it is convenient to work with a special case such as a rotation of coordinates through angle Φ around the z axis, as shown in the sketch. We have

$(\hat{\mathbf{i}}' \cdot \hat{\mathbf{i}}) = \cos\Phi$	$(\hat{\mathbf{j}}' \cdot \hat{\mathbf{i}}) = -\sin\Phi$	$(\hat{\mathbf{k}}' \cdot \hat{\mathbf{i}}) = 0$
$(\hat{\mathbf{i}}' \cdot \hat{\mathbf{j}}) = \sin\Phi$	$(\hat{\mathbf{j}}' \cdot \hat{\mathbf{j}}) = \cos\Phi$	$(\hat{\mathbf{k}}' \cdot \hat{\mathbf{j}}) = 0$
$(\hat{\mathbf{i}}' \cdot \hat{\mathbf{k}}) = 0$	$(\hat{\mathbf{j}}' \cdot \hat{\mathbf{k}}) = 0$	$(\hat{\mathbf{k}}' \cdot \hat{\mathbf{k}}) = 1.$

Hence the components of any vector $\mathbf{A} = (A_x,A_y,A_z)$ must transform according to the relations

$$A'_x = A_x \cos\Phi + A_y \sin\Phi$$
$$A'_y = -A_x \sin\Phi + A_y \cos\Phi \qquad 14.2$$
$$A'_z = A_z.$$



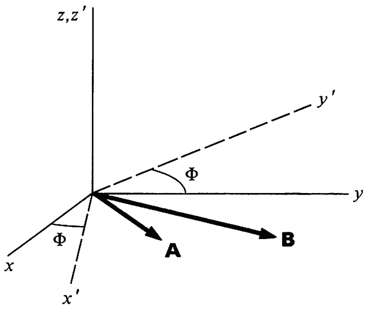
$x' = a + b = x\cos\Phi + y\sin\Phi$

For example, let $\mathbf{A} = \mathbf{r} = (x,y,z)$. Then

$$x' = x\cos\Phi + y\sin\Phi$$
$$y' = -x\sin\Phi + y\cos\Phi$$
$$z' = z.$$

These relations can be independently verified from the geometry. The drawing shows how the x' coordinate of point P is related to the coordinates (x,y).

Example 14.1 **Transformation Properties of the Vector Product**



In Chap. 1 we gave an essentially geometrical definition of the vector product. To demonstrate our new definition of a vector we shall prove that the components of the vector product transform as the components of a vector. For simplicity, we consider two coordinate systems, x, y, z and x', y', z', which differ by a rotation through angle Φ around the z axis, and two vectors **A** and **B** in the x, y plane. From the definition of vector product we have

$$\mathbf{C} = \mathbf{A}\times\mathbf{B} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ A_x & A_y & 0 \\ B_x & B_y & 0 \end{vmatrix} = \begin{vmatrix} \hat{\mathbf{i}}' & \hat{\mathbf{j}}' & \hat{\mathbf{k}}' \\ A'_x & A'_y & 0 \\ B'_x & B'_y & 0 \end{vmatrix}.$$

In the x, y, z system the components of **C** are

$$C_x = 0 \qquad 1a$$
$$C_y = 0 \qquad 1b$$
$$C_z = A_xB_y - A_yB_x \qquad 1c$$

and in the x', y', z' system they are

$$C'_x = 0 \qquad 2a$$
$$C'_y = 0 \qquad 2b$$
$$C'_z = A'_xB'_y - A'_yB'_x. \qquad 2c$$

If **C** is a vector, its components must obey the transformation law, Eq. (14.2):

$$C'_x = C_x\cos\Phi + C_y\sin\Phi \qquad 3a$$
$$C'_y = -C_x\sin\Phi + C_y\cos\Phi \qquad 3b$$
$$C'_z = C_z. \qquad 3c$$

Equations (3a) and (3b) are identically satisfied by Eqs. (1) and (2). To prove Eq. (3c), we need to show that $A'_xB'_y - A'_yB'_x = A_xB_y - A_yB_x$. From Eq. (14.2) we have

$$A'_x = A_x \cos\Phi + A_y \sin\Phi$$
$$A'_y = -A_x \sin\Phi + A_y \cos\Phi$$
$$B'_x = B_x \cos\Phi + B_y \sin\Phi$$
$$B'_y = -B_x \sin\Phi + B_y \cos\Phi.$$

Hence,

$$\begin{aligned} A'_xB'_y - A'_yB'_x &= (A_x \cos\Phi + A_y \sin\Phi)(-B_x \sin\Phi + B_y \cos\Phi) \\ &\quad - (-A_x \sin\Phi + A_y \cos\Phi)(B_x \cos\Phi + B_y \sin\Phi) \\ &= A_xB_y - A_yB_x \\ &= C_z. \end{aligned}$$

This proves that all three components of the vector product transform like the components of a vector so that the vector product is, in fact, a vector.

Example 14.2 **A Nonvector**

To give a counterexample to the cross product, suppose that we try to introduce a new type of vector multiplication, the vector "double cross" product $\mathbf{C} = \mathbf{A} \times\times \mathbf{B}$ defined by

$$C_x = A_yB_z + A_zB_y$$
$$C_y = A_zB_x + A_xB_z$$
$$C_z = A_xB_y + A_yB_x.$$

Is **C** actually a vector?

If we again take the case $\mathbf{A} = (A_x,A_y,0)$, $\mathbf{B} = (B_x,B_y,0)$, we have

$$C_x = 0$$
$$C_y = 0$$
$$C_z = A_xB_y + A_yB_x.$$

In the x', y', z' system the components are

$$C'_x = A'_yB'_z + A'_zB'_y = 0$$
$$C'_y = A'_zB'_x + A'_xB'_z = 0$$
$$C'_z = A'_xB'_y + A'_yB'_x.$$

The first two equations obey the transformation rule, Eqs. (3a) and (3b) of Example 14.1. However, when we evaluate the last equation we find that

$$\begin{aligned} C'_z &= (A_x \cos\Phi + A_y \sin\Phi)(-B_x \sin\Phi + B_y \cos\Phi) \\ &\quad + (-A_x \sin\Phi + A_y \cos\Phi)(B_x \cos\Phi + B_y \sin\Phi) \\ &= (A_xB_y + A_yB_x)(\cos^2\Phi - \sin^2\Phi) - 2(A_xB_x - A_yB_y)\cos\Phi \sin\Phi. \end{aligned}$$

It is apparent that $C_z' \neq C_z$, so that Eq. (3c) of Example 14.1 is not satisfied. The elements generated by the double cross product are not the components of a vector, and the double cross product is a useless operation.

Invariants of a Transformation

Any quantity which is unchanged by a general coordinate transformation is called an *invariant* of the transformation. Invariants play an important role in physics. They are the only entities suitable for the construction of physical laws, since the principle of relativity requires that the results of physical theories be independent of the choice of coordinate system (provided, of course, that the system is inertial).

We have so far encountered two classes of invariants—scalars and vectors. Scalars are single numbers and are unaffected by the choice of coordinates. Vectors are invariant under rotations of the coordinates by construction; we designed the transformation rule, Eq. (14.1), to assure this.

Any mathematical entity which is invariant under a rotation of coordinates is called a *tensor*. A scalar is a tensor of zeroth rank, and a vector is a tensor of the first rank. Tensors of higher rank also exist; the moment of inertia introduced in Chap. 7 is a tensor of the second rank.

The Transformation Properties of Physical Laws

We have used vector notation wherever possible because of its simplicity; one vector equation is easier to handle than three scalar equations. However, from the point of transformation theory, vectors have a deeper significance. Since we must be able to use any coordinate system we choose for describing physical events, it is essential that we be able to write physical laws in a form independent of coordinate systems. Thus, if an equation represents a statement of a physical law, both sides of the equation must transform the same way under a change of coordinates. For example, consider the equation for motion along some axis j: $F_j = ma_j$. Assuming m is a scalar, ma_j must be a component of a vector, since acceleration is a vector. Thus, F_j is a component of a vector along the same axis, and the general form of the equation must be $\mathbf{F} = m\mathbf{a}$. Once the law is in vector form, we can easily find the motion along any set of axes we choose. From this point of view, the vector nature of force, including the rule for superposition of forces, is a mathematical consequence of the requirement that the laws of motion be valid in all inertial systems.

The question arises as to whether the law of superposition of forces is a physical law or simply a mathematical result. It is, in fact, both. Experimentally, we find that the translation of a body can be described by only three independent equations, one for each coordinate axis; this implies that force has three independent components. According to transformation theory, the only three component entity suitable for describing physical laws is a vector, and vectors obey the law of superposition.

Scalar Invariants

We can use the dot product to combine two vectors to form a scalar. Since scalars are independent of the coordinate system, the dot product of two vectors is called a *scalar invariant*.

Let us show explicitly that the dot product $\mathbf{A} \cdot \mathbf{B}$ is a scalar invariant under rotations. Considering a rotation about the z axis for simplicity, we use Eq. (15.2) to obtain

$$\begin{aligned} A'_x B'_x + A'_y B'_y + A'_z B'_z = \\ &(A_x \cos\Phi + A_y \sin\Phi)(B_x \cos\Phi + B_y \sin\Phi) \\ &+ (-A_x \sin\Phi + A_y \cos\Phi) \\ &\quad (-B_x \sin\Phi + B_y \cos\Phi) + (A_z B_z) \\ &= A_x B_x + A_y B_y + A_z B_z. \end{aligned}$$

In particular, the dot product of a vector with itself, called the *norm* of the vector, is a scalar invariant:

$$A'^2_x + A'^2_y + A'^2_z = A_x{}^2 + A_y{}^2 + A_z{}^2.$$

The norm of the position vector $\mathbf{r}$ changes under a translation of coordinates but is invariant under pure rotations. We can use this to define a rotation of coordinates: it is any transformation which leaves $r^2 = x^2 + y^2 + z^2$ invariant.

14.3 Minkowski Space and Four-vectors

As we have discussed, it must be possible to express the laws of classical physics using entities like scalars and vectors, which are invariant under rotations of the coordinates x, y, z. From the mathematical point of view the Lorentz transformations have much in common with a spatial rotation: they are both linear transformations from one set of coordinates to another.

CHANGE OF COORDINATES UNDER A ROTATION	CHANGE OF COORDINATES UNDER LORENTZ TRANSFORMATION
$x' = x\cos\Phi + y\sin\Phi$	$x' = \gamma x - \gamma vt$
$y' = -x\sin\Phi + y\cos\Phi$	$y' = y$
$z' = z$	$z' = z$
$(t' = t)$	$t' = -(\gamma v/c^2)x + \gamma t$

Our object in this section is to find a way to write physical laws so that they are invariant under the Lorentz transformations. This assures that the laws will have the same form for observers in all inertial frames as required by the first postulate of relativity.

We shall start from the observation made in 1908 by the mathematician Minkowski that, with a slight change of notation, the Lorentz transformations represent a rotation in a four dimensional space. To introduce his line of reasoning, we return to the second postulate of relativity: the speed of light is the same for observers in all inertial frames. Consider two inertial systems x, y, z, t and x', y', z', t' moving with relative speed v in the x direction. If their origins coincide at $t = 0$ and a short light pulse is sent out from the origin at that instant, the locus of the pulse in the x, y, z, t system is $r = ct$, or

$$x^2 + y^2 + z^2 = (ct)^2,$$

while in the x', y', z', t' system it is

$$x'^2 + y'^2 + z'^2 = (ct')^2.$$

Comparing, we see that the quantity $x^2 + y^2 + z^2 - (ct)^2$ is equal to zero in each coordinate system; it appears to be a scalar invariant under the Lorentz transformations. We can show this directly by employing the Lorentz transformations, Eq. (11.3):

$$\begin{aligned} x'^2 + y'^2 + z'^2 - (ct')^2 &= \gamma^2(x - vt)^2 + y^2 + z^2 - \gamma^2c^2\left(t - \frac{vx}{c^2}\right)^2 \\ &= \frac{1}{1 - v^2/c^2}\left[x^2\left(1 - \frac{v^2}{c^2}\right) \right. \\ &\qquad \left. - c^2t^2\left(1 - \frac{v^2}{c^2}\right)\right] + y^2 + z^2 \\ &= x^2 + y^2 + z^2 - (ct)^2. \end{aligned} \qquad 14.3$$

In ordinary three dimensional space, the only transformation that leaves $x^2 + y^2 + z^2$ unchanged is a rotation. Minkowski considered a four dimensional space with coordinates x_1, x_2, x_3, x_4, where $x_1 = x$, $x_2 = y$, $x_3 = z$ and $x_4 = ict\,(i^2 = -1)$. With these coordinates,

$$x^2 + y^2 + z^2 - (ct)^2 = x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2$$

and Eq. (14.3) can be written

$$x_1'^2 + x_2'^2 + x_3'^2 + x_4'^2 = x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2.$$

It is apparent that $x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2$ is invariant under Lorentz transformations; by analogy with the three dimensional case, the Lorentz transformations represent a rotation of coordinates. The analogy also suggests that x_1, x_2, x_3, x_4 are the components of a true four dimensional vector.

The transformation rules for $(x_1,x_2,x_3,x_4) = (x,y,z,ict)$ are readily obtained from the Lorentz transformations.

$$\begin{aligned} x_1' &= \gamma(x_1 + i\beta x_4) \\ x_2' &= x_2 \\ x_3' &= x_3 \\ x_4' &= \gamma(x_4 - i\beta x_1), \end{aligned}$$

where $\beta = v/c$. (As usual, to simplify the algebra we restrict ourselves to systems whose relative motion is in the x direction.) It follows that any true four dimensional vector must transform in the same fashion. Such vectors are known as *four-vectors*. Thus the transformation rule for a four-vector $\underset{\rightarrow}{\mathbf{A}} = (A_1,A_2,A_3,A_4)$ is

$$\begin{aligned} A_1' &= \gamma(A_1 + i\beta A_4) \\ A_2' &= A_2 \\ A_3' &= A_3 \\ A_4' &= \gamma(A_4 - i\beta A_1). \end{aligned} \qquad 14.4$$

As we expect, the norm of $\underset{\rightarrow}{\mathbf{A}}$ is a Lorentz invariant.

$$A_1'^2 + A_2'^2 + A_3'^2 + A_4'^2 = A_1{}^2 + A_2{}^2 + A_3{}^2 + A_4{}^2.$$

The factor of c gives A_4 the same dimensions as the other components. From Eq. (14.4), we see that if A_1 is a real number, A_4 must be imaginary, as in the four-vector $\underset{\rightarrow}{\mathbf{s}} = (x,y,z,ict)$. The fact that the fourth component is imaginary arises from the essential difference between space and time.

In Minkowski's formulation of relativity, an event specified by x, y, z, t is viewed as a point x_1, x_2, x_3, x_4 in space-time. Minkowski called the four dimensional space-time manifold "world," although it has come to be called *Minkowski space*. A point in Minkowski space is called a *world point*. As a particle moves in space and time its successive world points trace out a *world line*.

The location of a world point is specified by its position four-vector

$$\vec{\mathbf{s}} = (x_1, x_2, x_3, x_4).$$

The Lorentz transformations, which relate an event in different coordinate systems, represent a transformation of the components of $\vec{\mathbf{s}}$ from one coordinate system to another.

The displacement between two world points is

$$\Delta\vec{\mathbf{s}} = (\Delta x,\ \Delta y,\ \Delta z,\ ic\ \Delta t)$$

or, in differential form,

$$d\vec{\mathbf{s}} = (dx,\ dy,\ dz,\ ic\ dt).$$

Since $d\vec{\mathbf{s}}$ is a four-vector, its norm is a Lorentz invariant. The norm is

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2\,dt^2.$$

A related Lorentz invariant that will be useful to us is $d\tau^2 = -ds^2/c^2$.

$$d\tau^2 = dt^2 - \frac{1}{c^2}(dx^2 + dy^2 + dz^2)$$

$d\tau$ has a simple interpretation. Consider a displacement $d\vec{\mathbf{s}}$ between two world points of a moving particle. In the rest frame of the particle, the space coordinates are constant, and therefore $dx = dy = dz = 0$. Thus $d\tau = dt$ in the rest frame; the world points are separated only in time. $d\tau$ is the time interval measured in the rest frame, and for this reason τ is known as the *proper time*.

Example 14.3 Time Dilation

We rederive the Einstein time dilation formula to show the power of four-vectors.

Consider an observer at rest in the x', y', z', t' system. In this system, the proper time interval between two world points is $d\tau = dt'$. In the x, y, z, t system moving with velocity $\vec{\mathbf{v}}$ relative to the first frame, the interval between the same points is given by

$$dt^2 - \frac{1}{c^2}(dx^2 + dy^2 + dz^2).$$

Since $d\tau^2$ is a Lorentz invariant, its value for the same world points is the same in all frames. Hence, we can equate its value in the rest frame to its value in the second frame.

$$dt'^2 = dt^2 - \frac{1}{c^2}(dx^2 + dy^2 + dz^2)$$

or

$$\left(\frac{dt'}{dt}\right)^2 = 1 - \frac{1}{c^2}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right].$$

Since $(dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2 = v^2$, we have

$$\left(\frac{dt'}{dt}\right)^2 = 1 - \frac{v^2}{c^2}$$

or

$$dt = \frac{dt'}{\sqrt{1 - v^2/c^2}} = \gamma\, d\tau.$$

In contrast to the derivation of Sec. 13.3, this treatment avoids hypothetical experiments and discussions of simultaneity.

Example 14.4 Construction of a Four-vector: The Four-velocity

In ordinary three dimensional space, dividing a vector by a scalar (a rotational invariant) yields another vector. Similarly, dividing a four-vector by a Lorentz invariant yields another four-vector.

Consider the displacement four-vector

$$d\underset{\rightarrow}{\mathbf{s}} = (dx, dy, dz, ic\, dt).$$

Dividing by the Lorentz invariant $d\tau$, we obtain a new four-vector

$$\frac{d\underset{\rightarrow}{\mathbf{s}}}{d\tau} = \left(\frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau}, ic\,\frac{dt}{d\tau}\right). \qquad 1$$

By analogy with the three dimensional case, we call $d\underset{\rightarrow}{\mathbf{s}}/d\tau$ the *four-velocity* $\underset{\rightarrow}{\mathbf{u}}$.

In the rest frame of the particle, $dx = dy = dz = 0$, and $d\tau = dt$. For a particle at rest

$$\underset{\rightarrow}{\mathbf{u}} = (0, 0, 0, ic). \qquad 2$$

The norm of $\underset{\rightarrow}{\mathbf{u}}$ is $(\underset{\rightarrow}{\mathbf{u}})^2 = -c^2$ and it has the same value in all frames. Obviously the four-velocity $\underset{\rightarrow}{\mathbf{u}}$ is very different physically from **u**, the familiar three dimensional velocity.

We now wish to find an expression for the four-velocity of a moving particle. Let the x, y, z, t system move with velocity $-\mathbf{u}$ relative to the rest frame of the particle. Using the time dilation formula of Example 14.3, we can write

$$dt = \gamma\, d\tau,$$

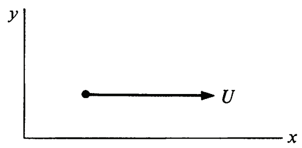
where dt is now the time interval in the moving frame. Using this in Eq. (1),

$$\begin{aligned} \underset{\rightarrow}{\mathbf{u}} &= \gamma\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}, ic\right) \\ &= \gamma(\mathbf{u}, ic), \qquad 3 \end{aligned}$$

where $\gamma = 1/\sqrt{1 - u^2/c^2}$.

We shall use $\underset{\rightarrow}{\mathbf{u}}$ in the next section to derive the momentum-energy four-vector. However, we shall first demonstrate how to transform a four-vector from one frame to another.

Example 14.5 **The Relativistic Addition of Velocities**

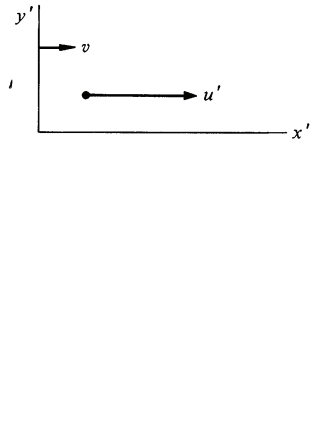


We can easily derive the formula for the relativistic addition of velocities by transforming the four-velocity $\underset{\rightarrow}{\mathbf{u}} = \gamma(\mathbf{u},ic)$ into successive frames with the aid of Eq. (14.4).

Consider a particle moving along the x direction of the x, y, z, t system with speed U. In this frame,

$$\underset{\rightarrow}{\mathbf{u}} = (u_1,u_2,u_3,u_4) = \Gamma(U,0,0,ic),$$

where $\Gamma = 1/\sqrt{1 - U^2/c^2}$. Consider a second frame x', y', z', t' moving along the x direction with speed v relative to the first frame. In this frame, the four-velocity of the particle is



$$\begin{aligned} \underset{\rightarrow}{\mathbf{u}}' &= (u_1',u_2',u_3',u_4') \\ &= \gamma'(\mathbf{u}',ic), \end{aligned}$$

where $\gamma' = 1/\sqrt{1 - u'^2/c^2}$. u' is the speed of the particle in the x', y', z', t' frame.

From the transformation rule, Eq. (14.4), and using $u_1 = \Gamma U$, $u_2 = 0$, $u_3 = 0$, $u_4 = i\Gamma c$,

$$\begin{aligned} u_1' &= \gamma(u_1 + i\beta u_4) = \gamma\Gamma(U - v) \\ u_2' &= u_2 = 0 \\ u_3' &= u_3 = 0 \\ u_4' &= \gamma(u_4 - i\beta u_1) = i\gamma\Gamma\left(c - \frac{vU}{c}\right) = ic\gamma\Gamma\left(1 - \frac{vU}{c^2}\right), \end{aligned}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$ and $\beta = v/c$. Hence,

$$\underset{\rightarrow}{\mathbf{u}}' = \gamma'(\mathbf{u}', ic)$$
$$= \gamma\Gamma\left[U - v, 0, 0, ic\left(1 - \frac{vU}{c^2}\right)\right].$$

Equating components,

$$\gamma' u' = \gamma\Gamma(U - v)$$

and

$$\gamma' = \gamma\Gamma\left(1 - \frac{vU}{c^2}\right).$$

Therefore,

$$u' = (\gamma\Gamma/\gamma')(U - v)$$
$$= \frac{U - v}{1 - vU/c^2},$$

which is Einstein's velocity addition formula for the case we are considering. The same procedure can be used to add nonparallel velocities.

14.4 The Momentum-energy Four-vector

In the last chapter we obtained expressions for the relativistic momentum and energy by rather labored arguments based on a hypothetical two body collision. In this section we shall obtain the same results in a much more direct manner by simply constructing a momentum-energy four-vector. We shall also obtain the relativistic expression for force, a difficult quantity to derive by the methods of the last chapter.

Our starting point is the observation that the classical momentum $m_0\mathbf{u}$ is not relativistically invariant since the classical velocity is not a four-vector. However, we found the form of the four-velocity $\underset{\rightarrow}{\mathbf{u}}$ in Example 14.4. Since the rest mass m_0 is a Lorentz invariant, the product $m_0\underset{\rightarrow}{\mathbf{u}}$ is a four-vector. It is natural to identify this with the relativistic momentum, and we therefore define the four-momentum

$$\underset{\rightarrow}{\mathbf{p}} = m_0\underset{\rightarrow}{\mathbf{u}}$$
$$= \gamma(m_0\mathbf{u}, im_0c)$$
$$= (m\mathbf{u}, imc)$$

or

$$\underset{\rightarrow}{\mathbf{p}} = (\mathbf{p}, imc). \qquad 14.5$$

Does the four-momentum obey a conservation law? Classically, the rate of change of momentum is equal to the applied force, so that the momentum of an isolated system is conserved. However, it is not obvious whether the four-momentum is similarly conserved since we have not developed a relativistic expression for force. Recall that we obtained the four-velocity by dividing $d\underset{\rightarrow}{\mathbf{s}}$ by the Lorentz invariant $d\tau$. Let us apply the same method to obtain the "time derivative" of $\underset{\rightarrow}{\mathbf{p}}$, and then *define* this equal to the four-force.

$$\underset{\rightarrow}{\mathbf{F}} = \frac{d\underset{\rightarrow}{\mathbf{p}}}{d\tau} = \left(\frac{d\mathbf{p}}{d\tau}, i\frac{d}{d\tau}mc\right) \qquad 14.6$$

$\underset{\rightarrow}{\mathbf{F}}$ is known as the *Minkowski force.*

If dt is the time interval in the observer's frame corresponding to the interval of proper time $d\tau$, then $dt = \gamma\, d\tau$ and we have

$$\underset{\rightarrow}{\mathbf{F}} = \gamma\left(\frac{d\mathbf{p}}{dt}, i\frac{d}{dt}mc\right).$$

In the classical limit, $d\mathbf{p}/dt = \mathbf{F}$. In order to conserve the momentum of an isolated system, we retain the identification of force with rate of change of momentum in all inertial systems. The Minkowski force becomes

$$\underset{\rightarrow}{\mathbf{F}} = \gamma\left(\mathbf{F}, i\frac{d}{dt}mc\right). \qquad 14.7$$

We have constructed $\underset{\rightarrow}{\mathbf{F}}$ so that four-momentum is conserved when the four-force is zero. Like all four-vectors, $\underset{\rightarrow}{\mathbf{F}}$ is relativistically invariant; if it is zero in one frame, it is zero in every frame. This assures us that if four-momentum is conserved in one inertial frame, it must be conserved in all inertial frames.

To interpret the fourth, or timelike component of $\underset{\rightarrow}{\mathbf{p}} = (\mathbf{p}, imc)$, we recall that classically $\mathbf{F}\cdot\mathbf{u}$ represents the rate at which work is done on a particle. By the work-energy theorem, $\mathbf{F}\cdot\mathbf{u} = dE/dt$, where E is the total energy. With this inspiration, let us examine $\underset{\rightarrow}{\mathbf{F}}\cdot\underset{\rightarrow}{\mathbf{u}}$ for a particle moving with velocity $\mathbf{u}$. Since $\underset{\rightarrow}{\mathbf{u}} = \gamma(\mathbf{u}, ic)$,

$$\underset{\rightarrow}{\mathbf{F}}\cdot\underset{\rightarrow}{\mathbf{u}} = \gamma^2\left(\mathbf{F}\cdot\mathbf{u} - \frac{d}{dt}mc^2\right).$$

Since the scalar product of two four-vectors is a Lorentz invariant, we are free to evaluate it in any frame we please. Let us evaluate $\underset{\rightarrow}{\mathbf{F}} \cdot \underset{\rightarrow}{\mathbf{u}}$ in the rest frame of the particle. In this frame, $(d\mathbf{p}/dt) \cdot \mathbf{u} = 0$ since $\mathbf{u} = 0$. We also have

$$\frac{d}{dt} mc^2 \bigg|_{u=0} = \frac{d}{dt}\left(\frac{m_0c^2}{\sqrt{1 - u^2/c^2}}\right)\bigg|_{u=0} = \frac{m_0u\ (du/dt)}{(1 - u^2/c^2)^{\frac{3}{2}}}\bigg|_{u=0} = 0.$$

Hence $\underset{\rightarrow}{\mathbf{F}} \cdot \underset{\rightarrow}{\mathbf{u}} = 0.$

$$\mathbf{F} \cdot \mathbf{u} - \frac{d}{dt} mc^2 = 0$$

or

$$\mathbf{F} \cdot \mathbf{u} = \frac{d}{dt} mc^2.$$

This relativistic result bears a close resemblance to the classical relation $\mathbf{F} \cdot \mathbf{u} = dE/dt$. We conclude that the relativistic equivalent of total energy is

$$E = mc^2.$$

The four-momentum becomes

$$\underset{\rightarrow}{\mathbf{p}} = (\mathbf{p}, imc) = \left(\mathbf{p}, \frac{iE}{c}\right). \qquad 14.8$$

$\underset{\rightarrow}{\mathbf{p}}$ is often called the *momentum-energy four-vector.*

We can generate a Lorentz invariant by taking the norm of $\underset{\rightarrow}{\mathbf{p}}$.

$$\underset{\rightarrow}{\mathbf{p}} \cdot \underset{\rightarrow}{\mathbf{p}} = p^2 - \frac{E^2}{c^2} = m_0{}^2\gamma^2(u^2 - c^2) = -m_0{}^2c^2.$$

Hence,

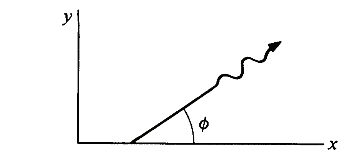
$$E^2 = p^2c^2 + (m_0c^2)^2,$$

a familiar result.

The Minkowski approach of generating four-vectors has led us in a natural way to relativistically correct expressions for momentum and energy. With this approach the conservation laws for energy and momentum appear as a single law: the conservation of four-momentum. In relativity, momentum and energy are different aspects of a single entity; this represents a significant simplification over classical physics, where the concepts are essentially unrelated.

We conclude this section with a few applications of the momentum-energy four-vector.

Example 14.6 The Doppler Effect, Once More

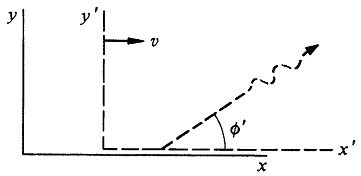


We have derived the relativistic expression for the Doppler effect by two different approaches: from a geometrical argument in Section 12.5 and by a dynamical argument in Example 13.8. In this example we obtain the same result by a third, much simpler, approach—four-vector invariance.

Consider a photon with energy $E = h\nu$ and momentum $h\nu/c$ traveling in the xy plane at angle ϕ with the x axis. The momentum in the x, y system is $\mathbf{p} = (h\nu/c)(\cos\phi, \sin\phi, 0)$. The momentum-energy four-vector is

$$\underset{\rightarrow}{\mathbf{p}} = \left(\mathbf{p}, \frac{iE}{c}\right)$$

$$= \frac{h\nu}{c}(\cos\phi, \sin\phi, 0, i).$$



In the x', y' system shown in the sketch, the four-momentum can be written

$$\underset{\rightarrow}{\mathbf{p}} = \frac{h\nu'}{c}(\cos\phi', \sin\phi', 0, i).$$

From Eq. (14.4) we have $p_4' = \gamma[p_4 - i(v/c)p_1]$. Hence,

$$i\frac{h\nu'}{c} = \gamma\left(i\frac{h\nu}{c} - i\frac{v}{c}\frac{h\nu}{c}\cos\phi\right)$$

$$\nu' = \gamma\nu\left(1 - \frac{v}{c}\cos\phi\right)$$

or

$$\nu = \frac{\nu'}{\gamma}\frac{1}{1 - (v/c)\cos\phi}$$

$$= \nu'\frac{\sqrt{1 - v^2/c^2}}{1 - (v/c)\cos\phi}$$

identical to our earlier result, Eq. (12.7).

Example 14.7 Relativistic Center of Mass Systems

The center of mass system we used in Chap. 4 to analyze collision problems is the coordinate system in which the spatial momentum is zero. In this example, we shall find the relativistic transformation from the laboratory system to the zero momentum frame.

Consider a collision between two particles with rest masses M_1 and M_2. Let particle 1 be moving with velocity **u** in the laboratory system and particle 2 be at rest. The momentum-energy four-vector of each particle is

$$\underset{\rightarrow}{\mathbf{p}}_1 = \left(p_1, 0, 0, \frac{iE_1}{c}\right)$$

$$\underset{\rightarrow}{\mathbf{p}}_2 = \left(0, 0, 0, \frac{iE_2}{c}\right).$$

The total momentum-energy is

$$\underset{\rightarrow}{\mathbf{P}} = \underset{\rightarrow}{\mathbf{p}}_1 + \underset{\rightarrow}{\mathbf{p}}_2 = \left(p_1, 0, 0, i\frac{E_1 + E_2}{c}\right). \qquad 1$$

In a frame moving along the x axis with speed V the spatial components of $\underset{\rightarrow}{\mathbf{P}}$ are, by Eq. (14.4),

$$P_1' = \Gamma\left(p_1 - V\frac{E_1 + E_2}{c^2}\right)$$

$$P_2' = 0 \qquad 2$$

$$P_3' = 0,$$

where $\Gamma = 1/\sqrt{1 - V^2/c^2}$.

In the center of mass system, $\mathbf{P}' = 0$. From Eq. (2) we see that the speed of this frame with respect to the laboratory frame is

$$V = \frac{p_1c^2}{E_1 + E_2}. \qquad 3$$

The energy available for physical processes such as the production of new particles or other inelastic events is the total energy in the center of mass system E'. In the center of mass frame, the momentum-energy four-vector is

$$\left(0, 0, 0, \frac{iE'}{c}\right). \qquad 4$$

We can find E' by using the invariance of the norm of $\underset{\rightarrow}{\mathbf{P}}$. From Eqs. (1) and (4),

$$-\frac{E'^2}{c^2} = p_1^2 - \frac{(E_1 + E_2)^2}{c^2}$$

or

$$E'^2 = (M_1c^2)^2 + 2E_1E_2 + E_2^2,$$

where we have used $p_1^2c^2 = E_1^2 - (M_1c^2)^2$. For our problem, $E_1 = \gamma M_1c^2$ and $E_2 = M_2c^2$, where $\gamma = 1/\sqrt{1 - u^2/c^2}$. Hence,

$$E' = (M_1^2 + M_2^2 + 2\gamma M_1M_2)^{\frac{1}{2}}c^2. \tag{5}$$

The total energy in the laboratory system is

$$E = (\gamma M_1 + M_2)c^2, \tag{6}$$

and the fraction of the initial energy available for physical processes is

$$\frac{E'}{E} = \frac{(M_1^2 + M_2^2 + 2\gamma M_1M_2)^{\frac{1}{2}}}{\gamma M_1 + M_2}. \tag{7}$$

An important practical case is that of equal masses $M_1 = M_2$. Equation (7) becomes

$$\begin{aligned}\frac{E'}{E} &= \frac{\sqrt{2}\sqrt{1+\gamma}}{1+\gamma}\\ &= \frac{\sqrt{2}}{\sqrt{1+\gamma}}.\end{aligned} \tag{8}$$

In the low velocity limit, $\gamma = 1$ and $E'/E = 1$. At low speeds, most of the energy is in rest mass energy and kinetic energy is relatively unimportant. To discuss the high-speed limit, it is convenient to write Eq. (8) in terms of the projectile energy $E_1 = \gamma Mc^2$.

$$\frac{E'}{E} = \frac{\sqrt{2}}{\sqrt{1 + E_1/Mc^2}}.$$

For $E_1 \gg Mc^2$, we have

$$\frac{E'}{E} \approx \frac{\sqrt{2Mc^2}}{\sqrt{E_1}}.$$

The useful fraction of energy decreases as $E_1^{-\frac{1}{2}}$. For example, the proton synchrotron at the National Accelerator Laboratory in Batavia, Illinois, can accelerate protons to an energy of 300 GeV (1 GeV $= 10^9$ eV). Since the rest mass of the proton is about 1 GeV, we see that for protons colliding with a hydrogen target, $E'/E \approx \sqrt{3}/\sqrt{200} \approx 0.1$. Only 30 GeV is available for interesting experiments.

By using identical beams colliding head on, the laboratory frame becomes the center of mass frame, and the total energy is available for inelastic events. This technique of colliding beams has been used extensively in electron accelerators and has proved feasible in proton machines as well.

Example 14.8 **Pair Production in Electron-electron Collisions**

In Example 13.7 we analyzed pair production, the process by which a photon collides with an electron to create an electron-positron pair. The threshold energy for the process was found to be $E = 2m_0c^2 = 1.02$ MeV, where $m_0c^2 = 0.51$ MeV is the rest energy of the electron or positron.

A related process is the production of an electron-positron pair by the collision of two electrons:

$$e^- + e^- \rightarrow e^- + e^- + (e^- + e^+).$$

The reaction evidently satisfies conservation of charge. The problem is to find the threshold energy for the process.

To describe the dynamics of the problem we introduce the following four-momenta:

$\underset{\rightarrow}{\mathbf{p}}_1$: electron 1 before the collision

$\underset{\rightarrow}{\mathbf{p}}_2$: electron 2 before the collision

$\underset{\rightarrow}{\mathbf{p}}_3$: electron 1 after the collision

$\underset{\rightarrow}{\mathbf{p}}_4$: electron 2 after the collision

$\underset{\rightarrow}{\mathbf{p}}_5$: electron created in e^--e^+ pair

$\underset{\rightarrow}{\mathbf{p}}_6$: positron created in e^--e^+ pair

Then conservation of four-momentum gives

$$\underset{\rightarrow}{\mathbf{p}}_1 + \underset{\rightarrow}{\mathbf{p}}_2 = \underset{\rightarrow}{\mathbf{p}}_3 + \underset{\rightarrow}{\mathbf{p}}_4 + \underset{\rightarrow}{\mathbf{p}}_5 + \underset{\rightarrow}{\mathbf{p}}_6.$$

Squaring, we have

$$(\underset{\rightarrow}{\mathbf{p}}_1 + \underset{\rightarrow}{\mathbf{p}}_2)^2 = (\underset{\rightarrow}{\mathbf{p}}_3 + \underset{\rightarrow}{\mathbf{p}}_4 + \underset{\rightarrow}{\mathbf{p}}_5 + \underset{\rightarrow}{\mathbf{p}}_6)^2. \tag{1}$$

Since each side of the equation is Lorentz invariant, we can compute the terms in whatever reference frame is most convenient.

Let us compute the left hand side of Eq. (1) in the laboratory frame. Taking particle 2 to be initially at rest, we have

$$\underset{\rightarrow}{\mathbf{p}}_1 = \left(\mathbf{p}_1, i\frac{E_1}{c}\right) \qquad \underset{\rightarrow}{\mathbf{p}}_2 = (0, im_0c)$$

and

$$\begin{aligned}(\underset{\rightarrow}{\mathbf{p}}_1 + \underset{\rightarrow}{\mathbf{p}}_2)^2 &= \underset{\rightarrow}{\mathbf{p}}_1{}^2 + \underset{\rightarrow}{\mathbf{p}}_2{}^2 + 2\underset{\rightarrow}{\mathbf{p}}_1 \cdot \underset{\rightarrow}{\mathbf{p}}_2 \\ &= -2(m_0c)^2 - 2m_0E_1, \end{aligned} \tag{2}$$

where we have used $\underset{\rightarrow}{\mathbf{p}}^2 = p^2 - E^2/c^2 = -m_0{}^2c^2$, valid for any particle.

The right hand side of Eq. (1) is most conveniently calculated in the center of mass frame. At threshold, all four particles are at rest. (This minimizes the energy and is consistent with the requirement that the total spatial momentum be zero in the center of mass frame.) Hence $\underset{\rightarrow}{\mathbf{p}}_3$, $\underset{\rightarrow}{\mathbf{p}}_4$, $\underset{\rightarrow}{\mathbf{p}}_5$, $\underset{\rightarrow}{\mathbf{p}}_6$ all have the form $(0,0,0,im_0c)$, and the right hand side of Eq. (1) becomes

$$(0,0,0,\ 4im_0c)^2 = -16(m_0c)^2. \qquad 3$$

Substituting Eqs. (2) and (3) in Eq. (1) gives

$$-2(m_0c)^2 - 2m_0E_1 = -16(m_0c)^2$$

or

$$E_1 = 7m_0c^2.$$

E_1 includes the rest energy of the projectile, so that the kinetic energy of the projectile at threshold is

$$\begin{aligned} K_1 &= E_1 - m_0c^2 \\ &= 6m_0c^2. \end{aligned}$$

The argument here can be applied to the production of other particles, for instance, to the production of a negative proton in the reaction

$$p^+ + p^+ \rightarrow p^+ + p^+ + (p^+ + p^-).$$

Since the proton rest mass is 0.94 GeV, the threshold kinetic energy for the production of negative protons is 6(0.94) GeV = 5.64 GeV. The Bevatron at the Lawrence Radiation Laboratory, Berkeley, California, was designed to accelerate protons to 6 GeV to allow this process to be observed. Owen Chamberlain and Emilio Segré received the Nobel Prize in 1959 for producing negative protons, or *antiprotons.*

14.5 Concluding Remarks

The special theory of relativity, far from representing a complete break with classical physics, has a heavy flavor of newtonian mechanics in its insistence on the equivalence of inertial frames. Essentially, Einstein generalized the work of Newton by bringing classical mechanics into accord with the requirements of electromagnetic theory.

Fundamentally, however, the emphases of special relativity are not the same as those of newtonian physics. Einstein's rejection of unobservable concepts like absolute space and time and his insistence on operational definitions related to observation were much more far-reaching than were Newton's efforts in this direction. Einstein laid the groundwork for the analysis of observables which was essential in the development of modern quantum mechanics. In addition, he made significant contributions to our philosophical understanding of how man obtains knowledge of the world.

As we have seen in this chapter, one of Einstein's great contributions was recognition of the power of transformation theory as an organizing principle in physics. Transformation theory unifies and simplifies the concepts of special relativity and has served as a knowledgeable guide in the search for new laws.

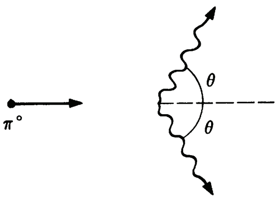
However, in spite of its power and harmony, special relativity is not a complete dynamical theory since it is inadequate to deal with accelerating reference frames. To Einstein this was a fundamental defect. According to Mach's principle of equivalence it is impossible to distinguish locally between an inertial system in a gravitational field and an accelerating coordinate system in free space. Therefore, by the principle of relativity, the frames must be equally valid for the description of physical phenomena. Since special relativity is incapable of dealing with accelerating reference frames, it is inherently incapable of dealing with gravitational fields.

Einstein went far toward removing these difficulties with his general theory of relativity, published in 1916. The general theory deals with transformations between all coordinate systems, not just inertial systems. It is essentially a theory of gravitation, since it is possible to "produce" a gravitational field merely by changing coordinate systems. From this point of view the effect of gravity is regarded as a local distortion in the geometry of space. Even in the gravitational field of the sun, however, effects attributable to general relativity are small and difficult to detect. For example, the deflection of starlight by the sun, one of the most dramatic effects predicted, amounts to only 1.7 seconds of arc.

General relativity's greatest impact has been on cosmology, since gravity is the only important force in the universe at large. Its role in terrestrial physics has been minor, however, partly because the effects are small and partly because so far it has not been extended to include electromagnetism. In contrast, special relativity has a multitude of applications and is part of the working knowledge of every physicist.

Einstein's impact on the twentieth century is difficult to assess in its entirety. He altered and enlarged our perceptions of the natural world, and in this respect he ranks among the great figures of Western thought.

Problems



14.1 A neutral pi meson, rest mass 135 MeV, decays symmetrically into two photons while moving at high speed. The energy of each photon in the laboratory system is 100 MeV.

a. Find the meson's speed V. Express your answer as a ratio V/c.

b. Find the angle θ in the laboratory system between the momentum of each photon and the initial line of motion.

Ans. $\theta \approx 42°$

14.2 A high energy photon (γ ray) collides with a proton at rest. A neutral pi meson (π°) is produced according to the reaction

$$\gamma + p \rightarrow p + \pi^\circ.$$

What is the minimum energy the γ ray must have for this reaction to occur? The rest mass of a proton is 938 MeV, and the rest mass of a π° is 135 MeV.

Ans. Approximately 145 MeV

14.3 A high energy photon (γ ray) hits an electron and produces an electron-positron pair according to the reaction

$$\gamma + e^- \rightarrow e^- + (e^- + e^+).$$

What is the minimum energy the γ ray must have for the reaction to occur?

14.4 A particle of rest mass M spontaneously decays from rest into two particles with rest masses m_1 and m_2. Show that the energies of the particles are

$$E_1 = (M^2 + m_1^2 - m_2^2)c^2/2M \qquad E_2 = (M^2 - m_1^2 + m_2^2)c^2/2M.$$

14.5 A nucleus of rest mass M_1 moving at high speed with kinetic energy K_1 collides with a nucleus of rest mass M_2 at rest. A nuclear reaction occurs according to the scheme

nucleus 1 + nucleus 2 $\rightarrow$ nucleus 3 + nucleus 4.

The rest masses of nuclei 3 and 4 are M_3 and M_4.

The rest masses are related by

$$(M_3 + M_4)c^2 = (M_1 + M_2)c^2 + Q,$$

where $Q > 0$. Find the minimum value of K_1 required to make the reaction occur, in terms of M_1, M_2, and Q.

Ans. clue. If $M_1 = M_2 = Q/c^2$, then $K_1 = 5Q/2$

14.6 A rocket of initial mass M_0 starts from rest and propels itself forward along the x axis by emitting photons backward.

a. Show that the four-momentum of the rocket's exhaust in the initial rest system can be written

$$\underset{\rightarrow}{\mathbf{p}} = \gamma M_f v(-1,0,0,i),$$

where M_f is the final mass of the rocket. (Note that this result is valid for the exhaust as a whole even though the photons are Doppler-shifted.)

b. Show that the final velocity of the rocket relative to the initial frame is

$$v = \frac{x^2 - 1}{x^2 + 1}c,$$

where x is the ratio of the rocket's initial mass to final mass, M_0/M_f.

14.7 Construct a four-vector representing acceleration. For simplicity, consider only straight line motion along the x axis. Let the instantaneous four-velocity be $\underset{\rightarrow}{\mathbf{u}}$.

Ans. clue. norm $= a^2/(1 - u^2/c^2)^3$, where $a = du/dt$

14.8 The function $f(x,t) = A \sin 2\pi[(x/\lambda) - \nu t]$ represents a sine wave of frequency ν and wavelength λ. The wave propagates along the x axis with velocity = wavelength $\times$ frequency $= \lambda\nu$. $f(x,t)$ can represent a light wave; A then corresponds to some component of the electromagnetic field which constitutes the light signal, and the wavelength and frequency satisfy $\lambda\nu = c$.

Consider the same wave in the coordinate system x', y', z', t' moving along the x axis at velocity v. In this reference frame the wave has the form

$$f'(x',t') = A' \sin 2\pi\left(\frac{x'}{\lambda'} - \nu' t'\right).$$

a. Show that the velocity of light is correctly given provided that $1/\lambda'$ and ν' are components of a four-vector k given in the x, y, z, t system by

$$\underset{\rightarrow}{\mathbf{k}} = 2\pi\left(\frac{1}{\lambda}, 0, 0, \frac{i\nu}{c}\right).$$

b. Using the result of part *a*, derive the result for the longitudinal Doppler shift by evaluating the frequency in a moving system.

c. Extend the analysis of part *b* to find the expression for the transverse Doppler shift by considering a wave propagating along the y axis.

INDEX