12 Electrodynamics and Relativity

12.1 ■ THE SPECIAL THEORY OF RELATIVITY

12.1.1 ■ Einstein's Postulates

Classical mechanics obeys the **principle of relativity**: the same laws apply in any **inertial reference frame**. By "inertial" I mean that the system is at rest or moving with constant velocity.[1] Imagine, for example, that you have loaded a billiard table onto a railroad car, and the train is going at constant speed down a smooth straight track. The game will proceed exactly the same as it would if the train were parked in the station; you don't have to "correct" your shots for the fact that the train is moving—indeed, if you pulled all the curtains, you would have no way of knowing whether the train was moving or not. Notice by contrast that you know *immediately* if the train speeds up, or slows down, or rounds a corner, or goes over a bump—the billiard balls roll in weird curved trajectories, and you yourself feel a lurch and spill coffee on your shirt. The laws of mechanics, then, are certainly *not* the same in *accelerating* reference frames.

In its application to classical mechanics, the principle of relativity is hardly new; it was stated clearly by Galileo. *Question:* does it also apply to the laws of electrodynamics? At first glance, the answer would seem to be *no*. After all, a charge in motion produces a magnetic field, whereas a charge at rest does not. A charge carried along by the train would generate a magnetic field, but someone on the train, applying the laws of electrodynamics in that system, would predict no magnetic field. In fact, many of the equations of electrodynamics, starting with the Lorentz force law, make explicit reference to "the" velocity of the charge. It certainly appears, therefore, that electromagnetic theory presupposes the existence of a unique stationary reference frame, with respect to which all velocities are to be measured.

And yet there is an extraordinary coincidence that gives us pause. Suppose we mount a wire loop on a freight car, and have the train pass between the poles of a

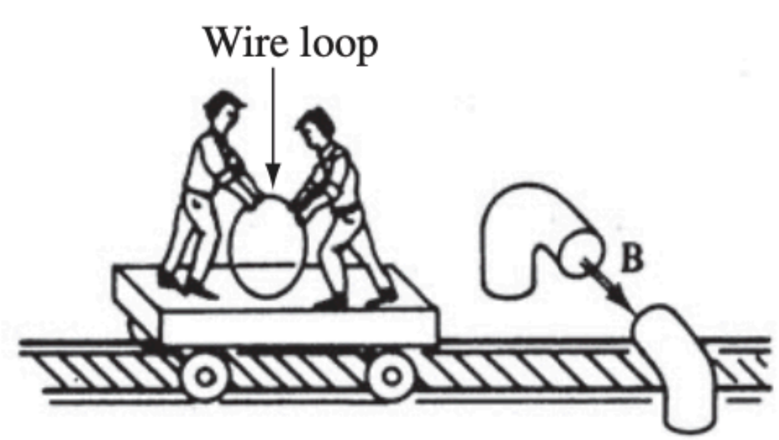

FIGURE 12.1

giant magnet (Fig. 12.1). As the loop rides through the magnetic field, a motional emf is established; according to the flux rule (Eq. 7.13),

$$\mathcal{E} = -\frac{d\Phi}{dt}.$$

This emf, remember, is due to the magnetic force on charges in the wire loop, which are moving along with the train. On the other hand, if someone on the train naïvely applied the laws of electrodynamics in *that* system, what would the prediction be? No *magnetic* force, because the loop is at rest. But as the magnet flies by, the magnetic field in the freight car changes, and a changing magnetic field induces an electric field, by Faraday's law. The resulting *electric* force would generate an emf in the loop given by Eq. 7.14:

$$\mathcal{E} = -\frac{d\Phi}{dt}.$$

Because Faraday's law and the flux rule predict exactly the same emf, people on the train will get the right answer, *even though their physical interpretation of the process is completely wrong*!

Or *is* it? Einstein could not believe this was a mere coincidence; he took it, rather, as a clue that electromagnetic phenomena, like mechanical ones, obey the principle of relativity. In his view, the analysis by the observer on the train is just as valid as that of the observer on the ground. If their *interpretations* differ (one calling the process electric, the other magnetic), so be it; their actual *predictions* are in agreement. Here's what he wrote on the first page of his 1905 paper introducing the **special theory of relativity**:

> It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends only on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either one or the other of these bodies is in motion. For if the magnet is in motion and the conductor at rest, there arises in the neighborhood of the magnet an electric field . . . producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighborhood of the magnet. In the conductor, however, we find an electromotive force . . . which gives rise—assuming equality of relative motion in the two cases discussed—to electric currents of the same path and intensity as those produced by the electric forces in the former case.
>
> Examples of this sort, together with unsuccessful attempts to discover any motion of the earth relative to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest.[2]

But I'm getting ahead of the story. To Einstein's predecessors, the equality of the two emfs was just a lucky accident; they had no doubt that one observer was right and the other was wrong. They thought of electric and magnetic fields as strains in an invisible jellylike medium called **ether**, which permeated all of space. The speed of the charge was to be measured *with respect to the ether*—only then would the laws of electrodynamics be valid. The train observer is wrong, because that frame is *moving* relative to the ether.

But wait a minute! How do we know the *ground* observer isn't moving relative to the ether, too? After all, the earth rotates on its axis once a day and revolves around the sun once a year; the solar system circulates around the galaxy, and for all I know the galaxy itself is moving at a high speed through the cosmos. All told, we should be traveling at well over 50 km/s with respect to the ether. Like a motorcycle rider on the open road, we face an "ether wind" of high velocity—unless by some miraculous coincidence we just happen to find ourselves in a tailwind of precisely the right strength, or the earth has some sort of "windshield" and drags its local supply of ether along with it. Suddenly it becomes a matter of crucial importance to *find* the ether frame, experimentally, or else *all* our calculations will be invalid.

The problem, then, is to determine our motion through the ether—to measure the speed and direction of the "ether wind." How shall we do it? At first glance you might suppose that practically *any* electromagnetic experiment would suffice: If Maxwell's equations are valid only with respect to the ether frame, any discrepancy between the experimental result and the theoretical prediction would be ascribable to the ether wind. Unfortunately, as nineteenth-century physicists soon realized, the anticipated error in a typical experiment is extremely small; as in the example above, "coincidences" always seem to conspire to hide the fact that we are using the "wrong" reference frame. So it takes an uncommonly delicate experiment to do the job.

Now, among the results of classical electrodynamics is the prediction that electromagnetic waves travel through the vacuum at a speed

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

relative (presumably) *to the ether*. In principle, then, one should be able to detect the ether wind by simply measuring the speed of light in various directions. Like a motorboat on a river, the net speed "downstream" should be a maximum, for here the light is swept along by the ether; in the opposite direction, where it is bucking the current, the speed should be a minimum (Fig. 12.2).

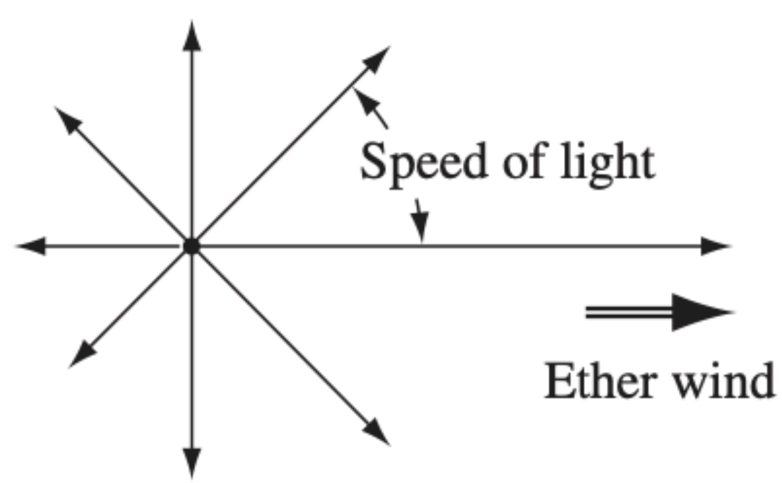

FIGURE 12.2

Nowadays, when students are taught in high school to snicker at the naïveté of the ether model, it takes some imagination to comprehend how utterly perplexing this result must have been at the time. All other waves (water waves, sound waves, waves on a string) travel at a prescribed speed *relative to the propagating medium* (the stuff that does the waving), and if this medium is in motion with respect to the observer, the net speed is always greater "downstream" than "upstream." Over the next 20 years, a series of improbable schemes were concocted in an effort to explain why this does *not* occur with light. Michelson and Morley themselves interpreted their experiment as confirmation of the "ether drag" hypothesis, which held that the earth somehow pulls the ether along with it. But this was found to be inconsistent with other observations, notably the aberration of starlight.[3] Various so-called "emission" theories were proposed, according to which the speed of electromagnetic waves is governed by the motion of the *source*—as it would be in a corpuscular theory (conceiving of light as a stream of particles). Such theories called for implausible modifications in Maxwell's equations, but in any event they were discredited by experiments using extraterrestrial light sources. Meanwhile, Fitzgerald and Lorentz suggested that the ether wind physically compresses all matter (including the Michelson-Morley apparatus itself) in just the right way to compensate for, and thereby conceal, the variation in speed with direction. As it turns out, there is a grain of truth in this, although their idea of the *reason* for the contraction was quite wrong.

At any rate, it was not until Einstein that anyone took the Michelson-Morley result at face value, and suggested that the speed of light is a universal constant, the same in all directions, regardless of the motion of the observer or the source. There *is* no ether wind because there is no ether. *Any* inertial system is a suitable reference frame for the application of Maxwell's equations, and the velocity of a charge is to be measured *not* with respect to a (nonexistent) absolute rest frame, nor with respect to a (nonexistent) ether, but simply with respect to the particular inertial system you happen to have chosen.

Inspired, then, both by internal theoretical hints (the fact that the laws of electrodynamics are such as to give the right answer even when applied in the "wrong" system) and by external empirical evidence (the Michelson-Morley experiment[4]), Einstein proposed his two famous postulates:

1. **The principle of relativity.** The laws of physics apply in all inertial reference systems.
2. **The universal speed of light.** The speed of light in vacuum is the same for all inertial observers, regardless of the motion of the source.

The special theory of relativity derives from these two postulates. The first elevates Galileo's observation about classical mechanics to the status of a general law, applying to *all* of physics. It states that there is no absolute rest system. The second might be considered Einstein's response to the Michelson-Morley experiment. It means that there is no ether. (Some authors consider Einstein's second postulate redundant—no more than a special case of the first. They maintain that the very existence of ether would violate the principle of relativity, in the sense that it would define a unique stationary reference frame. I think this is nonsense. The existence of air as a medium for sound does not invalidate the theory of relativity. Ether is no more an absolute rest system than the water in a goldfish bowl—which is a *special* system, if you happen to be the goldfish, but scarcely "absolute.")[5]

Unlike the principle of relativity, which had roots going back several centuries, the universal speed of light was radically new—and, on the face of it, preposterous. For if I walk 5 mi/h down the corridor of a train going 60 mi/h, my net speed relative to the ground is "obviously" 65 mi/h—the speed of A (me) with respect to C (ground) is equal to the speed of A relative to B (train) plus the speed of B relative to C:

$$v_{AC} = v_{AB} + v_{BC}. \tag{12.1}$$

And yet, if A is a *light* signal (whether it comes from a flashlight on the train or a lamp on the ground or a star in the sky) Einstein would have us believe that its speed is c relative to the train *and* c relative to the ground:

$$v_{AC} = v_{AB} = c. \tag{12.2}$$

Clearly, Eq. 12.1, which we now call **Galileo's velocity addition rule** (no one before Einstein would have bothered to give it a name at all) is incompatible with the second postulate. In special relativity, as we shall see, it is replaced by **Einstein's velocity addition rule**:

$$\boxed{v_{AC} = \frac{v_{AB} + v_{BC}}{1 + (v_{AB}v_{BC}/c^2)}.} \tag{12.3}$$

For "ordinary" speeds ($v_{AB} \ll c$, $v_{BC} \ll c$), the denominator is so close to 1 that the discrepancy between Galileo's formula and Einstein's formula is negligible. On the other hand, Einstein's formula has the desired property that if $v_{AB} = c$, then *automatically* $v_{AC} = c$:

$$v_{AC} = \frac{c + v_{BC}}{1 + (c v_{BC}/c^2)} = c.$$

But how can Galileo's rule, which seems to rely on nothing but common sense, possibly be wrong? And if it *is* wrong, what does this do to all of classical physics? The answer is that special relativity compels us to alter our notions of space and time themselves, and therefore also of such derived quantities as velocity, momentum, and energy. Although it developed historically out of Einstein's contemplation of electrodynamics, the special theory is not limited to any particular class of phenomena—rather, it is a description of the space-time "arena" in which *all* physical phenomena take place. And in spite of the reference to the speed of light in the second postulate, relativity has nothing to do with light: c is a fundamental velocity, and it happens that light travels at that speed, but it is perfectly possible to conceive of a universe in which there are no electric charges, and hence no electromagnetic fields or waves, and yet relativity would still prevail. Because relativity defines the structure of space and time, it claims authority not merely over all presently known phenomena, but over those not yet discovered. It is, as Kant would say, a "prolegomenon to any future physics."

12.1.2 ■ The Geometry of Relativity

In this section I present a series of *gedanken* (thought) experiments that serve to introduce the three most striking geometrical consequences of Einstein's postulates: time dilation, Lorentz contraction, and the relativity of simultaneity. In Sect. 12.1.3 the same results will be derived more systematically, using Lorentz transformations.

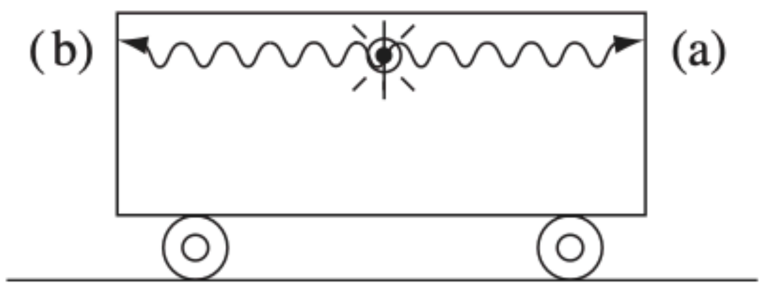

FIGURE 12.4

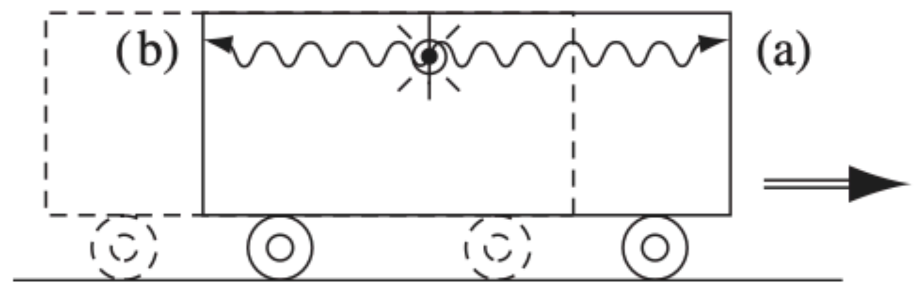

FIGURE 12.5

(i) The relativity of simultaneity. Imagine a freight car, traveling at constant speed along a smooth, straight track (Fig. 12.4). In the very center of the car there hangs a light bulb. When someone switches it on, the light spreads out in all directions at speed c. Because the lamp is equidistant from the two ends, an observer on the train will find that the light reaches the front end at the same instant as it reaches the back end: The two events in question—(a) light reaches the front end (and maybe a buzzer goes off) and (b) light reaches the back end (another buzzer sounds)—occur *simultaneously*.

However, to an observer on the *ground* these same two events are *not* simultaneous. For as the light travels out from the bulb (going at speed c in both directions—that's the second postulate), the train itself moves forward, so the beam going to the back end has a shorter distance to travel than the one going forward (Fig. 12.5). According to this observer, therefore, event (b) happens *before* event (a). An observer passing by on an express train, meanwhile, would report that (a) preceded (b). *Conclusion*:

Two events that are simultaneous in one inertial system are not, in general, simultaneous in another.

Naturally, the train has to be going awfully fast before the discrepancy becomes detectable—that's why you don't notice it all the time.

Of course, it's always possible for a naïve witness to be *mistaken* about simultaneity: a person sitting in the back corner of the car would *hear* buzzer *b* before buzzer *a*, simply because he's closer to the source of the sound, and a child might infer that *b* actually *rang* before *a*. But this is a trivial error, having nothing to do with special relativity—*obviously*, you must correct for the time the signal (sound, light, carrier pigeon, or whatever) takes to reach you. When I speak of an **observer**, I mean someone with the sense to make this correction, and an **observation** is what he records *after* doing so. What you hear or see, therefore, is not the same as what you *observe*. An observation is an artificial reconstruction after the fact, when all the data are in, and it doesn't depend on *where* the observer is located. In fact, a wise observer will avoid the whole problem by stationing assistants at strategic locations, each equipped with a watch synchronized to a master clock, so that time measurements can be made right at the scene. I belabor this point in order to emphasize that the relativity of simultaneity is a genuine discrepancy between measurements made by competent observers in relative motion, not a simple mistake arising from a failure to account for the travel time of light signals.

(ii) Time dilation. Now let's consider a light ray that leaves the bulb and strikes the floor of the car directly below. *Question*: How long does it take the light to make this trip? From the point of view of an observer on the train, the answer is easy: If the height of the car is h, the time is

$$\Delta\bar{t} = \frac{h}{c}. \tag{12.4}$$

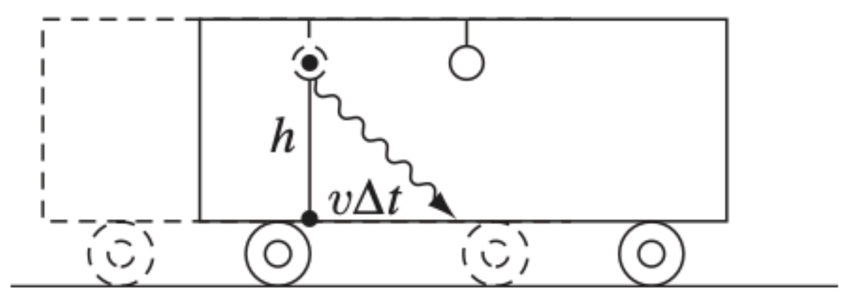

FIGURE 12.7

(I'll use an overbar to denote measurements made on the train.) On the other hand, as observed from the ground, this same ray must travel farther, because the train itself is moving. From Fig. 12.7, I see that this distance is $\sqrt{h^2 + (v\Delta t)^2}$, so

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

Solving for Δt, we have

$$\Delta t = \frac{h}{c}\frac{1}{\sqrt{1 - v^2/c^2}},$$

and therefore

$$\boxed{\Delta\bar{t} = \sqrt{1 - v^2/c^2}\,\Delta t.} \tag{12.5}$$

Evidently the time elapsed between the *same two events*—(a) light leaves bulb, and (b) light strikes center of floor—is different for the two observers. In fact, the interval recorded on the train clock, $\Delta\bar{t}$, is *shorter* by the factor

$$\boxed{\gamma \equiv \frac{1}{\sqrt{1 - v^2/c^2}}.} \tag{12.6}$$

Conclusion:

Moving clocks run slow.

This is called **time dilation**. It doesn't have anything to do with the mechanics of clocks; it's a statement about the nature of time, which applies to *all* properly functioning timepieces.

Of all Einstein's predictions, none has received more spectacular and persuasive confirmation than time dilation. Most elementary particles are unstable: they disintegrate after a characteristic lifetime[6] that varies from one species to the next. The lifetime of a neutron is 15 min; of a muon, 2×10^{-6} s; and of a neutral pion, 9×10^{-17} s. But these are lifetimes of particles at *rest*. When particles are moving at speeds close to c they last much longer, for their internal clocks (whatever it is that tells them when their time is up) are running slow, in accordance with Einstein's time dilation formula.

Example 12.1. A muon is traveling through the laboratory at three-fifths the speed of light. How long does it last?

Solution

In this case,

$$\gamma = \frac{1}{\sqrt{1-(3/5)^2}} = \frac{5}{4},$$

so it lives longer (than at rest) by a factor of $\frac{5}{4}$:

$$\frac{5}{4} \times (2 \times 10^{-6})\text{ s} = 2.5 \times 10^{-6}\text{ s}.$$

It may strike you that time dilation is inconsistent with the principle of relativity. For if the ground observer says the train clock runs slow, the train observer can with equal justice claim that the *ground* clock runs slow—after all, from the train's point of view it is the ground that is in motion. Who's right? *Answer:* They're *both* right! On closer inspection, the "contradiction," which seems so stark, evaporates. Let me explain: In order to check the rate of the train clock, the ground observer uses *two* of his own clocks (Fig. 12.8): one to compare times at the beginning of the interval, when the train clock passes point A, the other to compare times at the end of the interval, when the train clock passes point B. Of course, he must be careful to synchronize his clocks before the experiment. What he finds is that while the train clock ticked off, say, 3 minutes, the interval between his own two clock readings was 5 minutes. He concludes that the train clock runs slow.

Meanwhile, the observer on the train is checking the rate of the ground clock by the same procedure: She uses two carefully synchronized train clocks, and compares times with a single ground clock as it passes by each of them in turn (Fig. 12.9). She finds that while the ground clock ticks off 3 minutes, the interval between her train clocks is 5 minutes, and concludes that the *ground* clock runs slow. Is there a contradiction? *No*, for the two observers have measured *different things*. The ground observer compared *one* train clock with *two* ground clocks; the train observer compared one *ground* clock with two *train* clocks. Each followed a sensible and correct procedure, comparing a single moving clock with two stationary ones. "So what," you say, "the stationary clocks were synchronized in each instance, so it cannot matter that they used two different ones." But there's the rub: *Clocks that are properly synchronized in one system will not*

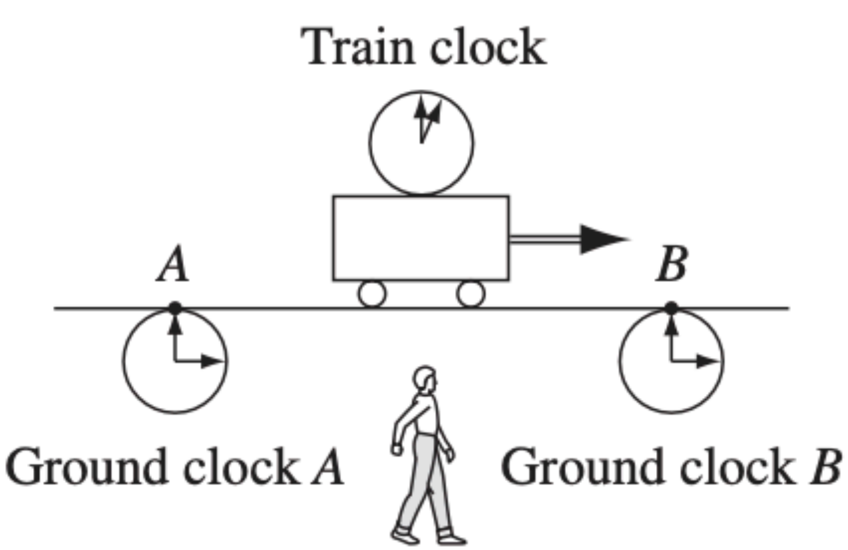

FIGURE 12.8

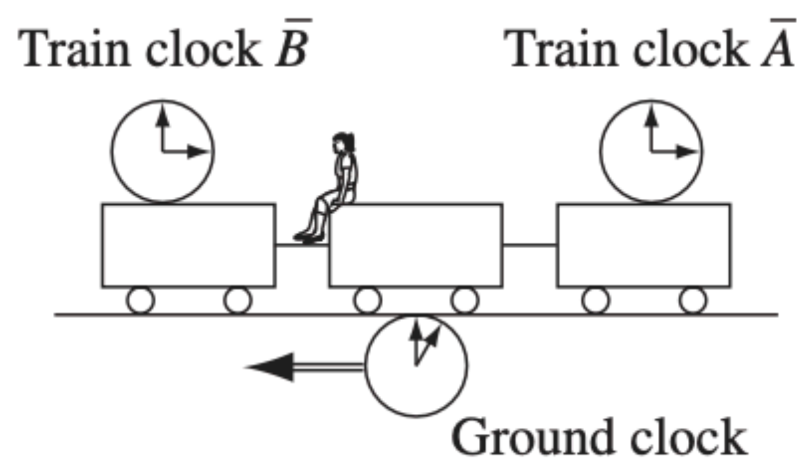

FIGURE 12.9

be synchronized when observed from another system. They *can't* be, for to say that two clocks are synchronized is to say that they read 12 noon *simultaneously*, and we have already learned that what's simultaneous to one observer is *not* simultaneous to another. So whereas each observer conducted a perfectly sound measurement, from his/her own point of view, the *other* observer (watching the process) considers that she/he made the most elementary blunder in the book, by using two unsynchronized clocks. That's how, in spite of the fact that *his* clocks "actually" run slow, he manages to conclude that *hers* are running slow (and vice versa).

Because moving clocks are not synchronized, it is essential when checking time dilation to focus attention on a *single* moving clock. *All* moving clocks run slow by the same factor, but you can't start timing on one clock and then switch to another because they weren't in step to begin with. But you can use as many *stationary* clocks (stationary with respect to you, the observer) as you please, for they *are* properly synchronized (moving observers would dispute this, but that's *their* problem).

Example 12.2. The twin paradox. On her 21st birthday, an astronaut takes off in a rocket ship at a speed of $\frac{12}{13}c$. After 5 years have elapsed on her watch, she turns around and heads back at the same speed to rejoin her twin brother, who stayed at home. *Question*: How old is each twin at their reunion?

Solution

The traveling twin has aged 10 years (5 years out, 5 years back); she arrives at home just in time to celebrate her 31st birthday. However, as viewed from earth, the moving clock has been running slow by a factor

$$\gamma = \frac{1}{\sqrt{1-(12/13)^2}} = \frac{13}{5}.$$

The time elapsed on earthbound clocks is $\frac{13}{5} \times 10 = 26$, and her brother will be therefore celebrating his 47th birthday—he is now 16 years older than his twin sister! But don't be deceived: This is no fountain of youth for the traveling twin, for though she may die later than her brother, she will not have lived any *more*—she's just done it *slower*. During the flight, all her biological processes—metabolism, pulse, thought, and speech—are subject to the same time dilation that affects her watch.

The so-called **twin paradox** arises when you try to tell this story from the point of view of the *traveling* twin. She sees the earth fly off at $\frac{12}{13}c$, turn around after 5 years, and return. From her point of view, it would seem, *she's* at rest, whereas her *brother* is in motion, and hence it is *he* who should be younger at the reunion. An enormous amount has been written about the twin paradox, but the truth is there's really no paradox here at all: this second analysis is simply wrong. The two twins are not equivalent. The traveling twin experiences *acceleration* when she turns around to head home, but her brother does not. To put it in fancier language, the traveling twin is not in an inertial system—more precisely, she's in one inertial system on the way out and a completely different inertial system on the way back. You'll see in Prob. 12.16 how to analyze this problem correctly from her perspective, but as far as the resolution of the "paradox" is concerned, it is enough to note that the *traveling twin cannot claim to be a stationary observer* because you can't undergo acceleration and remain stationary.

(iii) Lorentz contraction. For the third gedanken experiment you must imagine that we have set up a lamp at one end of a boxcar and a mirror at the other, so that a light signal can be sent down and back (Fig. 12.10). *Question*: How long does the signal take to complete the round trip? To an observer on the train, the answer is

$$\Delta\bar{t} = 2\frac{\Delta\bar{x}}{c}, \tag{12.7}$$

where $\Delta\bar{x}$ is the length of the car (the overbar, as before, denotes measurements made on the train). To an observer on the ground, the process is more complicated because of the motion of the train. If Δt_1 is the time for the light signal to reach the front end, and Δt_2 is the return time, then (see Fig. 12.11):

$$\Delta t_1 = \frac{\Delta x + v\Delta t_1}{c}, \quad \Delta t_2 = \frac{\Delta x - v\Delta t_2}{c},$$

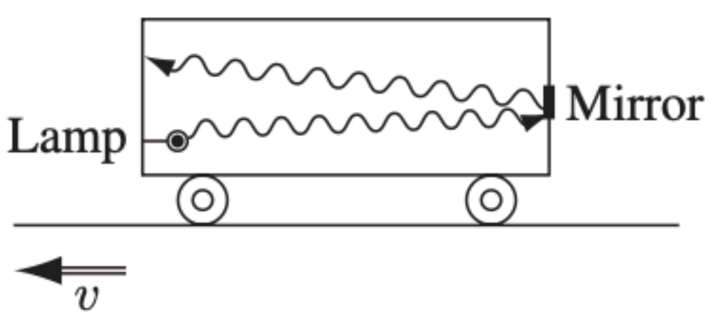

FIGURE 12.10

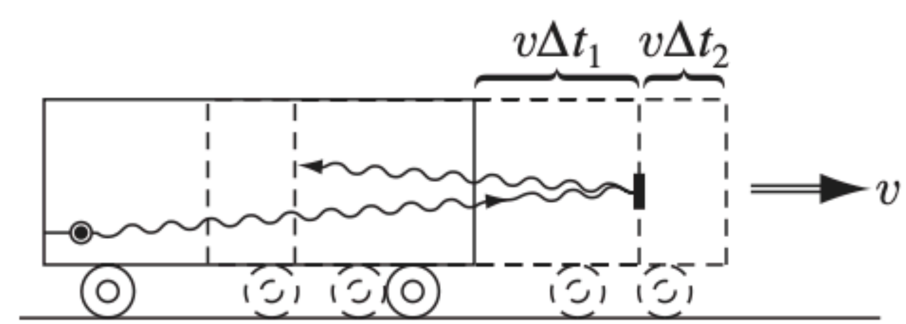

FIGURE 12.11

or, solving for Δt_1 and Δt_2:

$$\Delta t_1 = \frac{\Delta x}{c - v}, \quad \Delta t_2 = \frac{\Delta x}{c + v}.$$

So the round-trip time is

$$\Delta t = \Delta t_1 + \Delta t_2 = 2\frac{\Delta x}{c}\frac{1}{(1 - v^2/c^2)}. \tag{12.8}$$

But these intervals are related by the time dilation formula, Eq. 12.5:

$$\Delta\bar{t} = \sqrt{1 - v^2/c^2}\,\Delta t.$$

Applying this to Eqs. 12.7 and 12.8, I conclude that

$$\boxed{\Delta\bar{x} = \frac{1}{\sqrt{1 - v^2/c^2}}\Delta x.} \tag{12.9}$$

The length of the boxcar is not the same when measured by an observer on the ground, as it is when measured by an observer on the train—from the ground point of view, it is somewhat *shorter*. *Conclusion*:

Moving objects are shortened.

We call this **Lorentz contraction**. Notice that the same factor,

$$\gamma \equiv \frac{1}{\sqrt{1 - v^2/c^2}},$$

appears in both the time dilation formula and the Lorentz contraction formula. This makes it all very easy to remember: Moving clocks run slow, moving sticks are shortened, and the factor is always γ.

Of course, the observer on the train doesn't think her car is shortened—her meter sticks are contracted by that same factor, so all her measurements come out the same as when the train was standing in the station. In fact, from *her* point of view it is objects on the *ground* that are shortened. This raises again a paradoxical problem: If A says B's sticks are short, and B says A's sticks are short, who is right? *Answer:* They *both* are! But to reconcile the rival claims we must study carefully the actual process by which length is measured.

Suppose you want to find the length of a board. If it's at rest (with respect to you) you simply lay your ruler down next to the board, record the readings at each end, and subtract (Fig. 12.12). (If you're really clever, you'll line up the left end of the ruler against the left end of the board—then you only have to read *one* number.)

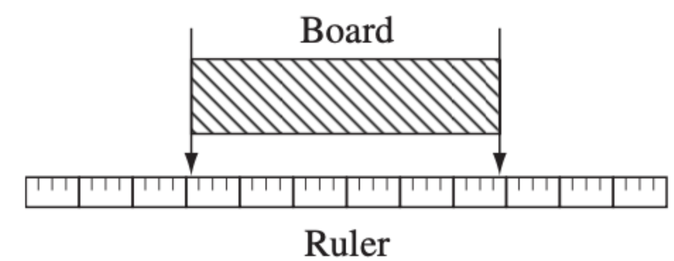

FIGURE 12.12

But what if the board is *moving*? Same story, only this time, of course, you must be careful to read the two ends *at the same instant of time.* If you don't, the board will move in the course of measurement, and obviously you'll get the wrong answer. But therein lies the problem: Because of the relativity of simultaneity the two observers disagree on what constitutes "the same instant of time." When the person on the ground measures the length of the boxcar, he reads the position of the two ends at the same instant *in his system*. But the person on the train, watching him do it, complains that he read the front end first, then waited a moment before reading the back end. *Naturally*, he came out short, in spite of the fact that (to her) he was using an undersized meter stick, which would otherwise have yielded a number too *large*. Both observers measure lengths correctly (from the point of view of their respective inertial frames), and each finds the other's sticks to be shortened. Yet there is no inconsistency, for they are measuring different things, and each considers the other's method improper.

Example 12.3. The barn and ladder paradox. Unlike time dilation, there is no direct experimental confirmation of Lorentz contraction, simply because it's too difficult to get an object of measurable size going anywhere near the speed of light. The following parable illustrates how bizarre the world would be if the speed of light were more accessible.

There once was a farmer who had a ladder too long to store in his barn (Fig. 12.13a). He chanced one day to read some relativity, and a solution to his problem suggested itself. He instructed his daughter to run with the ladder as fast as she could—the moving ladder having Lorentz-contracted to a size the barn could easily accommodate, she was to rush through the door, whereupon the

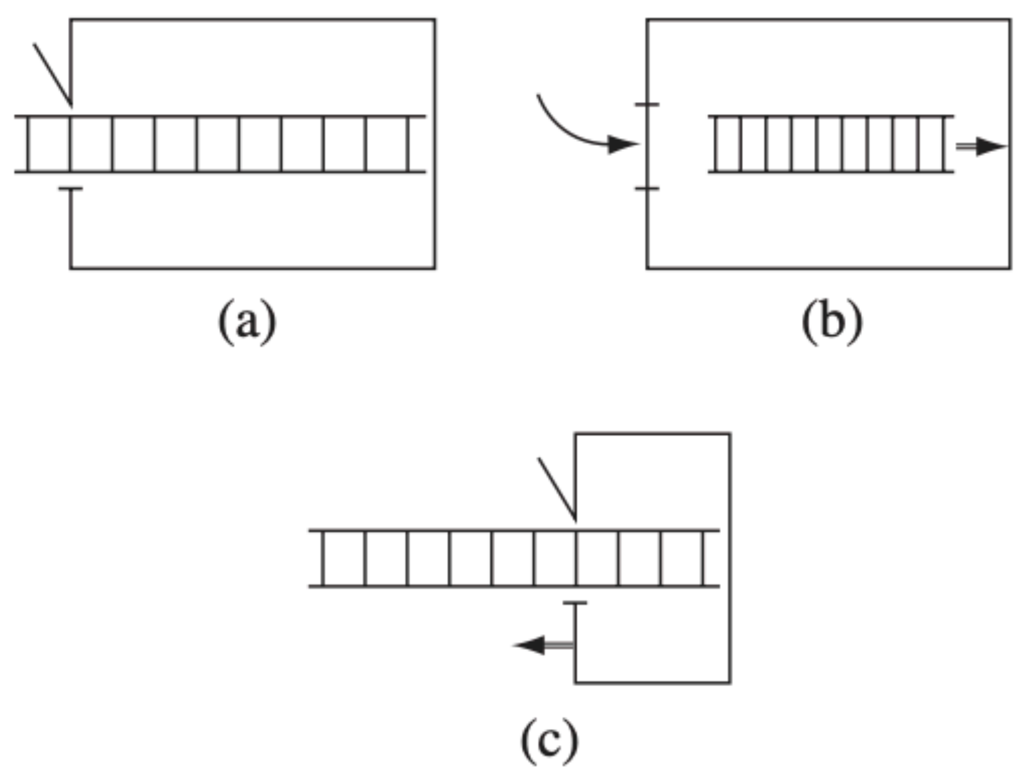

FIGURE 12.13

Solution

They're *both* right! When you say "the ladder is in the barn," you mean that all parts of it are inside *at one instant of time*, but in view of the relativity of simultaneity, that's a condition that depends on the observer. There are really *two* relevant events here:

a. Back end of ladder makes it in the door.
b. Front end of ladder hits far wall of barn.

The farmer says a occurs before b, so there *is* a time when the ladder is entirely within the barn; his daughter says b precedes a, so there is *not*. *Contradiction?* Nope—just a difference in perspective.

"But *come* now," I hear you protest, "when it's all over and the dust clears, either the ladder is inside the barn, or it isn't. There can be no dispute about *that*." Quite so, but now you're introducing a new element into the story: What happens *as the ladder is brought to a stop?* Suppose the farmer grabs the last rung of the ladder firmly with one hand, while he slams the door with the other. Assuming it remains intact, the ladder must now stretch out to its rest length. Evidently, the front end keeps going, even after the rear end has stopped! Expanding like an accordian, the front end of the ladder smashes into the far side of the barn. In truth, the whole notion of a "rigid" object loses its meaning in relativity, for when it changes its speed, different parts do not in general accelerate simultaneously—in this way, the material stretches or shrinks to reach the length appropriate to its new velocity.[7]

But to return to the question at hand: When the ladder finally comes to a stop, is it inside the barn or not? The answer is indeterminate. When the front end of the ladder hits the far side of the barn, something has to give, and the farmer is left either with a broken ladder inside the barn or with the ladder intact poking through a hole in the wall. In any event, he is unlikely to be pleased with the outcome.

One final comment on Lorentz contraction. A moving object is shortened *only along the direction of its motion:*

Dimensions perpendicular to the velocity are not contracted.

Indeed, in deriving the time dilation formula I took it for granted that the *height* of the train is the same for both observers. I'll now justify this, using a lovely gedanken experiment suggested by Taylor and Wheeler.[8] Imagine that we build a wall beside the railroad tracks, and 1 m above the rails (*as measured on the ground*), we paint a horizontal blue line. When the train goes by, a passenger leans out the window holding a wet paintbrush 1 m above the rails, *as measured on the train*, leaving a horizontal *red* line on the wall. *Question*: Does the passenger's red line lie above or below our blue one? If the rule were that perpendicular directions contract, then the person on the ground would predict that the *red* line is lower, while the person on the train would say it's the *blue* one (to the latter, of course, the *ground* is moving). The principle of relativity says that both observers are equally justified, but they cannot both be right. No subtleties of simultaneity or synchronization can rationalize this contradiction; either the blue line is higher or the red one is—*unless they exactly coincide*, which is the inescapable conclusion. There *cannot* be a law of contraction (or expansion) of perpendicular dimensions, for it would lead to irreconcilably inconsistent predictions.

12.1.3 ■ The Lorentz Transformations

Any physical process consists of one or more **events**. An "event" is something that takes place at a specific location (x, y, z), at a precise time (t). The explosion of a firecracker, for example, is an event; a tour of Europe is not. Suppose we know the coordinates (x, y, z, t) of a particular event E in *one* inertial system $\mathcal{S}$, and we would like to calculate the coordinates $(\bar{x}, \bar{y}, \bar{z}, \bar{t})$ of that *same event* in some other inertial system $\bar{\mathcal{S}}$. What we need is a "dictionary" for translating from the language of $\mathcal{S}$ to the language of $\bar{\mathcal{S}}$.

We may as well orient our axes as shown in Fig. 12.16, so that $\bar{\mathcal{S}}$ slides along the x axis at speed v. If we "start the clock" $(t = 0)$ at the moment the origins ($\mathcal{O}$ and $\bar{\mathcal{O}}$) coincide, then at time t, $\bar{\mathcal{O}}$ will be a distance vt from $\mathcal{O}$, and hence

$$x = d + vt, \tag{12.10}$$

where d is the distance from $\bar{\mathcal{O}}$ to $\bar{A}$ at time t ($\bar{A}$ is the point on the $\bar{x}$ axis that is even with E when the event occurs). Before Einstein, anyone would have said immediately that

$$d = \bar{x}, \tag{12.11}$$

and thus constructed the "dictionary"

$$\left.\begin{array}{ll} \text{(i)} & \bar{x} = x - vt, \\ \text{(ii)} & \bar{y} = y, \\ \text{(iii)} & \bar{z} = z, \\ \text{(iv)} & \bar{t} = t. \end{array}\right\} \tag{12.12}$$

These are now called the **Galilean transformations**, though they scarcely deserve so fine a title—the last one, in particular, went without saying, since everyone assumed the flow of time is the same for all observers. In the context of special

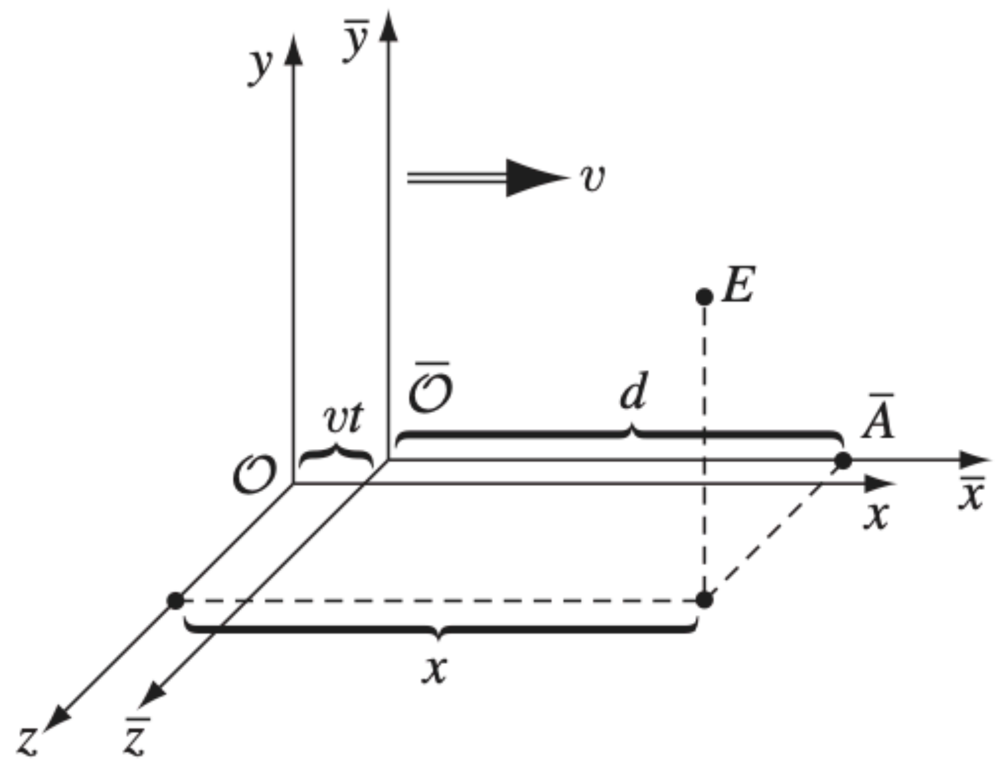

FIGURE 12.16

relativity, however, we must expect (iv) to be replaced by a rule that incorporates time dilation, the relativity of simultaneity, and the nonsynchronization of moving clocks. Likewise, there will be a modification in (i) to account for Lorentz contraction. As for (ii) and (iii), they, at least, remain unchanged, for we have already seen that there can be no modification of lengths perpendicular to the motion.

But where does the classical derivation of (i) break down? *Answer:* In Eq. 12.11. For d is the distance from $\bar{\mathcal{O}}$ to $\bar{A}$ *as measured in* $\mathcal{S}$, whereas $\bar{x}$ is the distance from $\bar{\mathcal{O}}$ to $\bar{A}$ *as measured in* $\bar{\mathcal{S}}$. Because $\bar{\mathcal{O}}$ and $\bar{A}$ are at rest in $\bar{\mathcal{S}}$, $\bar{x}$ is the "moving stick," which appears contracted to $\mathcal{S}$:

$$d = \frac{1}{\gamma}\bar{x}. \tag{12.13}$$

When this is inserted in Eq. 12.10 we obtain the relativistic version of (i):

$$\bar{x} = \gamma(x - vt). \tag{12.14}$$

Of course, we could have run the same argument from the point of view of $\bar{\mathcal{S}}$. The diagram (Fig. 12.17) looks similar, but in this case it depicts the scene *at time* $\bar{t}$, whereas Fig. 12.16 showed the scene *at time* t. (Note that t and $\bar{t}$ represent the same physical instant *at* E, but not elsewhere, because of the relativity of simultaneity.) If we assume that $\bar{\mathcal{S}}$ also starts its clock when the origins coincide, then at time $\bar{t}$, $\mathcal{O}$ will be a distance $v\bar{t}$ from $\bar{\mathcal{O}}$, and therefore

$$\bar{x} = \bar{d} - v\bar{t}, \tag{12.15}$$

where $\bar{d}$ is the distance from $\mathcal{O}$ to A at time $\bar{t}$, and A is that point on the x axis that is even with E when the event occurs. The classical physicist would have said that $x = \bar{d}$, and, using (iv), recovered (i). But, as before, relativity demands that we observe a subtle distinction: x is the distance from $\mathcal{O}$ to A *in* $\mathcal{S}$, whereas $\bar{d}$ is the distance from $\mathcal{O}$ to A *in* $\bar{\mathcal{S}}$. Because $\mathcal{O}$ and A are at rest in $\mathcal{S}$, x is the "moving stick," and

$$\bar{d} = \frac{1}{\gamma} x. \tag{12.16}$$

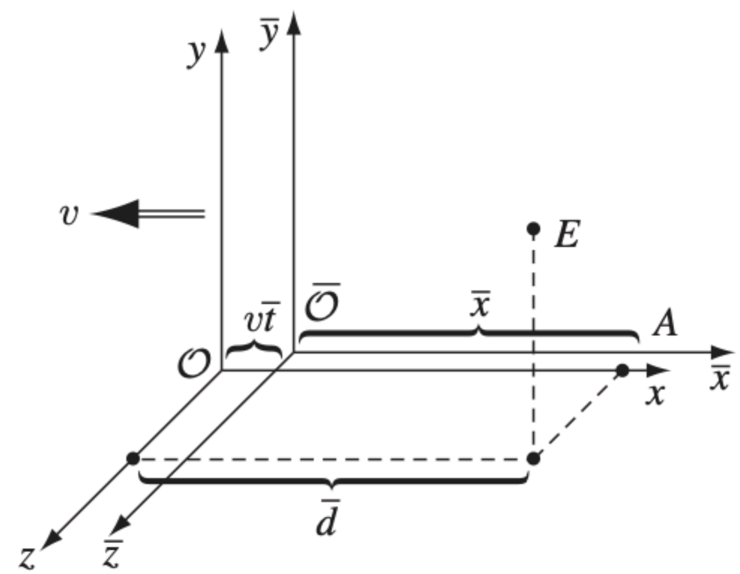

FIGURE 12.17

It follows that

$$x = \gamma(\bar{x} + v\bar{t}). \tag{12.17}$$

This last equation comes as no surprise, for the symmetry of the situation dictates that the formula for x, in terms of $\bar{x}$ and $\bar{t}$, should be identical to the formula for $\bar{x}$ in terms of x and t (Eq. 12.14), except for a switch in the sign of v. (If $\bar{\mathcal{S}}$ is going to the *right* at speed v, with respect to $\mathcal{S}$, then $\mathcal{S}$ is going to the *left* at speed v, with respect to $\bar{\mathcal{S}}$.) Nevertheless, this is a useful result, for if we substitute $\bar{x}$ from Eq. 12.14, and solve for $\bar{t}$, we complete the relativistic "dictionary":

$$\boxed{\begin{aligned} &\text{(i)}\ \ \bar{x} = \gamma(x - vt), \\ &\text{(ii)}\ \ \bar{y} = y, \\ &\text{(iii)}\ \ \bar{z} = z, \\ &\text{(iv)}\ \ \bar{t} = \gamma\left(t - \frac{v}{c^2}x\right). \end{aligned}} \tag{12.18}$$

These are the famous **Lorentz transformations**, with which Einstein replaced the Galilean ones. They contain all the geometrical information in the special theory, as the following examples illustrate. The reverse dictionary, which carries you from $\bar{\mathcal{S}}$ back to $\mathcal{S}$, can be obtained algebraically by solving (i) and (iv) for x and t, or, more simply, by switching the sign of v:

$$\left.\begin{aligned} &\text{(i}'\text{)}\quad x = \gamma(\bar{x} + v\bar{t}), \\ &\text{(ii}'\text{)}\quad y = \bar{y}, \\ &\text{(iii}'\text{)}\quad z = \bar{z}, \\ &\text{(iv}'\text{)}\quad t = \gamma\left(\bar{t} + \frac{v}{c^2}\bar{x}\right). \end{aligned}\right\} \tag{12.19}$$

Example 12.4. Simultaneity, synchronization, and time dilation. Suppose event A occurs at $x_A = 0, t_A = 0$, and event B occurs at $x_B = b, t_B = 0$. The two events are simultaneous in $\mathcal{S}$ (they both take place at $t = 0$). But they are *not* simultaneous in $\bar{\mathcal{S}}$, for the Lorentz transformations give $\bar{x}_A = 0, \bar{t}_A = 0$ and $\bar{x}_B = \gamma b, \bar{t}_B = -\gamma(v/c^2)b$. According to the $\bar{\mathcal{S}}$ clocks, then, B occurred *before* A. This is nothing *new*, of course—just the relativity of simultaneity. But I wanted you to see how it follows from the Lorentz transformations.

Suppose that at time $t = 0$ observer $\mathcal{S}$ decides to examine *all* the clocks in $\bar{\mathcal{S}}$. He finds that they read *different* times, depending on their location; from (iv):

$$\bar{t} = -\gamma\frac{v}{c^2}x.$$

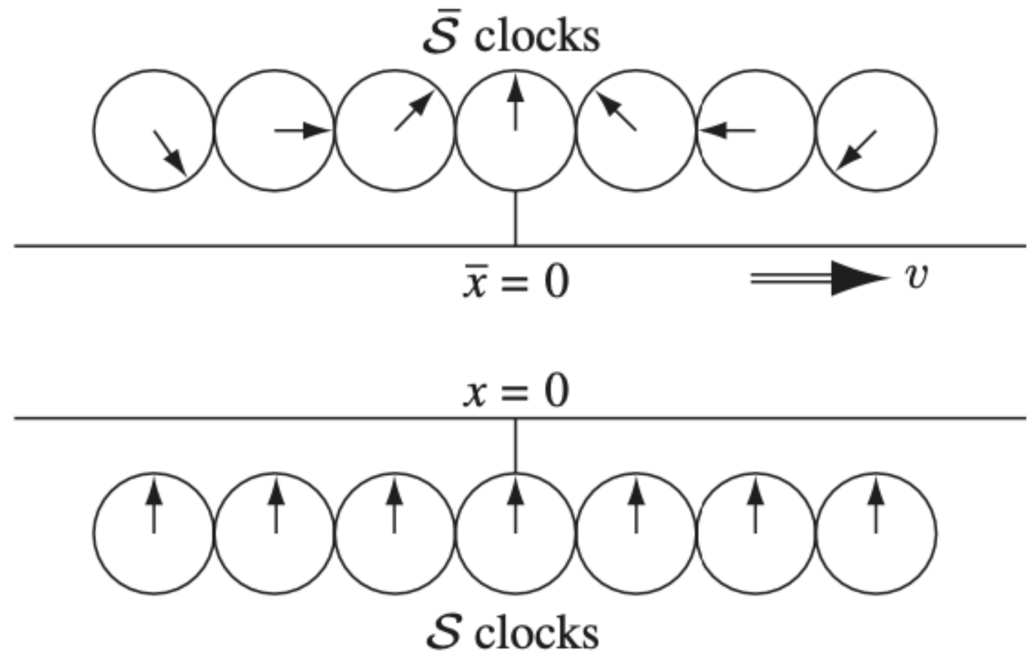

FIGURE 12.18

Those to the left of the origin (negative x) are *ahead*, and those to the right are *behind*, by an amount that increases in proportion to their distance (Fig. 12.18). Only the master clock at the origin reads $\bar{t} = 0$. Thus, the nonsynchronization of moving clocks, too, follows directly from the Lorentz transformations. Of course, from the $\bar{S}$ viewpoint it is the S clocks that are out of synchronization, as you can check by putting $\bar{t} = 0$ into equation (iv′).

Finally, suppose S focuses his attention on a single clock at rest in the $\bar{S}$ frame (say, the one at $\bar{x} = a$), and watches it over some interval Δt. How much time elapses on the moving clock? Because $\bar{x}$ is fixed, (iv′) gives $\Delta t = \gamma \Delta \bar{t}$, or

$$\Delta \bar{t} = \frac{1}{\gamma} \Delta t.$$

That's the old time dilation formula, derived now from the Lorentz transformations. Please note that it's $\bar{x}$ we hold fixed, here, because we're watching *one moving clock*. If you hold x fixed, then you're watching a whole series of different $\bar{S}$ clocks as they pass by, and that won't tell you whether any one of them is running slow.

Example 12.5. Lorentz contraction. Imagine a stick at rest in $\bar{\mathcal{S}}$ (hence moving to the right at speed v in $\mathcal{S}$). Its rest length (that is, its length as measured in $\bar{\mathcal{S}}$) is $\Delta\bar{x} = \bar{x}_r - \bar{x}_l$, where the subscripts denote the right and left ends of the stick. If an observer in $\mathcal{S}$ were to measure the stick, he would subtract the positions of the two ends at one instant of *his* time t: $\Delta x = x_r - x_l$ (for $t_l = t_r$). According to (i), then,

$$\Delta x = \frac{1}{\gamma}\Delta\bar{x}.$$

This is the old Lorentz contraction formula. Note that it's t we hold fixed, here, because we're talking about a measurement made by $\mathcal{S}$, and he marks off the two ends at the same instant of his time. ($\bar{\mathcal{S}}$ doesn't have to be so fussy, since the stick is at rest in her frame.)

Example 12.6. Einstein's velocity addition rule. Suppose a particle moves a distance dx (in $\mathcal{S}$) in a time dt. Its velocity u is then

$$u = \frac{dx}{dt}.$$

In $\bar{\mathcal{S}}$, meanwhile, it has moved a distance

$$d\bar{x} = \gamma(dx - v\,dt),$$

as we see from (i), in a time given by (iv):

$$d\bar{t} = \gamma\left(dt - \frac{v}{c^2}dx\right).$$

The velocity in $\bar{\mathcal{S}}$ is therefore

$$\bar{u} = \frac{d\bar{x}}{d\bar{t}} = \frac{\gamma(dx - v\,dt)}{\gamma\left(dt - v/c^2 dx\right)} = \frac{(dx/dt - v)}{1 - v/c^2 dx/dt} = \frac{u - v}{1 - uv/c^2}. \qquad (12.20)$$

This is **Einstein's velocity addition rule**. To recover the more transparent notation of Eq. 12.3, let A be the particle, B be $\mathcal{S}$, and C be $\bar{\mathcal{S}}$; then $u = v_{AB}$, $\bar{u} = v_{AC}$, and $v = v_{CB} = -v_{BC}$, so Eq. 12.20 becomes

$$v_{AC} = \frac{v_{AB} + v_{BC}}{1 + (v_{AB}v_{BC}/c^2)}.$$

12.1.4 ■ The Structure of Spacetime

(i) Four-vectors. The Lorentz transformations take on a simpler appearance when expressed in terms of the quantities

$$x^0 \equiv ct, \quad \beta \equiv \frac{v}{c}. \tag{12.21}$$

Using x^0 (instead of t) and β (instead of v) amounts to changing the unit of time from the *second* to the *meter*—1 meter of x^0 corresponds to the time it takes light to travel 1 meter (in vacuum). If, at the same time, we number the x, y, z coordinates, so that

$$x^1 = x, \quad x^2 = y, \quad x^3 = z, \tag{12.22}$$

then the Lorentz transformations read

$$\left.\begin{aligned} \bar{x}^0 &= \gamma(x^0 - \beta x^1), \\ \bar{x}^1 &= \gamma(x^1 - \beta x^0), \\ \bar{x}^2 &= x^2, \\ \bar{x}^3 &= x^3. \end{aligned}\right\} \tag{12.23}$$

Or, in matrix form:

$$\begin{pmatrix} \bar{x}^0 \\ \bar{x}^1 \\ \bar{x}^2 \\ \bar{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}. \tag{12.24}$$

Letting Greek indices run from 0 to 3, this can be distilled into a single equation:

$$\bar{x}^\mu = \sum_{\nu=0}^{3} (\Lambda^\mu_\nu) x^\nu, \tag{12.25}$$

where Λ is the **Lorentz transformation matrix** in Eq. 12.24 (the superscript μ labels the row, the subscript ν labels the column). One virtue of writing things in this abstract manner is that we can handle in the same format a more general transformation, in which the relative motion is *not* along a common $x\bar{x}$ axis; the matrix Λ would be more complicated, but the structure of Eq. 12.25 is unchanged.

If this reminds you of the *rotations* we studied in Sect. 1.1.5, it's no accident. There we were concerned with the change in components when you switch to a *rotated* coordinate system; here we are interested in the change of components when you go to a *moving* system. In Chapter 1 we defined a (3-)vector as any set of three components that transform under rotations the same way (x, y, z) do; by extension, we now define a **4-vector** as any set of *four* components that transform in the same manner as (x^0, x^1, x^2, x^3) under Lorentz transformations:

$$\bar{a}^\mu = \sum_{\nu=0}^{3} \Lambda^\mu_\nu a^\nu. \tag{12.26}$$

For the particular case of a transformation along the x axis,

$$\left.\begin{aligned} \bar{a}^0 &= \gamma(a^0 - \beta a^1), \\ \bar{a}^1 &= \gamma(a^1 - \beta a^0), \\ \bar{a}^2 &= a^2, \\ \bar{a}^3 &= a^3. \end{aligned}\right\} \tag{12.27}$$

There is a 4-vector analog to the dot product ($\mathbf{A} \cdot \mathbf{B} \equiv A_x B_x + A_y B_y + A_z B_z$), but it's not just the sum of the products of like components; rather, the zeroth components have a minus sign:

$$-a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3. \tag{12.28}$$

This is the **four-dimensional scalar product**; you should check for yourself (Prob. 12.17) that it has the same value in all inertial systems:

$$-\bar{a}^0\bar{b}^0 + \bar{a}^1\bar{b}^1 + \bar{a}^2\bar{b}^2 + \bar{a}^3\bar{b}^3 = -a^0b^0 + a^1b^1 + a^2b^2 + a^3b^3; \tag{12.29}$$

just as the ordinary dot product is **invariant** (unchanged) under rotations, this combination is invariant under Lorentz transformations.

To keep track of the minus sign, it is convenient to introduce the **covariant** vector a_μ, which differs from the **contravariant** a^μ only in the sign of the zeroth component:

$$a_\mu = (a_0, a_1, a_2, a_3) \equiv (-a^0, a^1, a^2, a^3). \tag{12.30}$$

You must be scrupulously careful about the placement of indices in this business: *upper* indices designate *contravariant* vectors; *lower* indices are for *covariant* vectors. Raising or lowering the temporal index costs a minus sign ($a_0 = -a^0$); raising or lowering a spatial index changes nothing ($a_1 = a^1$, $a_2 = a^2$, $a_3 = a^3$). Formally,

$$a_\mu = \sum_{\nu=0}^{3} g_{\mu\nu} a^\nu, \quad \text{where} \quad g_{\mu\nu} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{12.31}$$

is the (**Minkowski**) **metric**.[10]

The scalar product can now be written with the summation symbol,

$$\sum_{\mu=0}^{3} a^{\mu} b_{\mu},$$

or, more compactly still,

$$a^{\mu} b_{\mu}. \tag{12.32}$$

(Summation is *implied* whenever a Greek index is repeated in a product—once as a covariant index and once as contravariant. This is called the **Einstein summation convention**, after its inventor, who regarded it as one of his most important contributions.) Of course, we could just as well take care of the minus sign by switching to covariant b:

$$a_{\mu} b^{\mu} = a^{\mu} b_{\mu} = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3. \tag{12.33}$$

(ii) The invariant interval. The scalar product of a 4-vector with *itself*, $a^{\mu} a_{\mu} = -(a^0)^2 + (a^1)^2 + (a^2)^2 + (a^3)^2$, can be positive (if the "spatial" terms dominate) or negative (if the "temporal" term dominates) or zero:

If $a^{\mu} a_{\mu} > 0$, a^{μ} is called **spacelike**.

If $a^{\mu} a_{\mu} < 0$, a^{μ} is called **timelike**.

If $a^{\mu} a_{\mu} = 0$, a^{μ} is called **lightlike**.

Suppose event A occurs at $(x^0_A, x^1_A, x^2_A, x^3_A)$, and event B at $(x^0_B, x^1_B, x^2_B, x^3_B)$. The difference,

$$\Delta x^\mu \equiv x^\mu_A - x^\mu_B, \tag{12.35}$$

is the **displacement 4-vector**. The scalar product of Δx^μ with itself is called the **invariant interval** between two events:

$$I \equiv (\Delta x)^\mu (\Delta x)_\mu = -(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 = -c^2 t^2 + d^2, \tag{12.36}$$

where t is the time difference between the two events and d is their spatial separation. When you transform to a moving system, the *time* between A and B is altered ($\bar{t} \neq t$), and so is the *spatial separation* ($\bar{d} \neq d$), but the interval I remains the same.

If the displacement between two events is timelike ($I < 0$), there exists an inertial system (accessible by Lorentz transformation) in which they occur at the same point. For if I hop on a train going from (A) to (B) at the speed $v = d/t$, leaving event A when it occurs, I shall be just in time to pass B when *it* occurs; in the train system, A and B take place at the same point. You cannot do this for a *spacelike* interval, of course, because v would have to be greater than c, and no observer can exceed the speed of light (γ would be imaginary and the Lorentz transformations would be nonsense). On the other hand, if the displacement is spacelike ($I > 0$), then there exists a system in which the two events occur at the same time (see Prob. 12.21). And if the displacement is lightlike ($I = 0$), then the two events could be connected by a light signal.

(iii) Space-time diagrams. If you want to represent the motion of a particle graphically, the normal practice is to plot the position versus time (that is, x runs vertically and t horizontally). On such a graph, the velocity can be read off as the slope of the curve. For some reason, the convention is reversed in relativity: everyone plots position horizontally and time (or, better, $x^0 = ct$) vertically. Velocity is then given by the *reciprocal* of the slope. A particle at rest is represented by a vertical line; a photon, traveling at the speed of light, is described by a 45° line; and a rocket going at some intermediate speed follows a line of slope $c/v = 1/\beta$ (Fig. 12.21). We call such plots **Minkowski diagrams**.

The trajectory of a particle on a Minkowski diagram is called a **world line**. Suppose you set out from the origin at time $t = 0$. Because no material object can travel faster than light, your world line can never have a slope less than 1. Accordingly, your motion is restricted to the wedge-shaped region bounded by the two 45° lines (Fig. 12.22). We call this your "future," in the sense that it is the

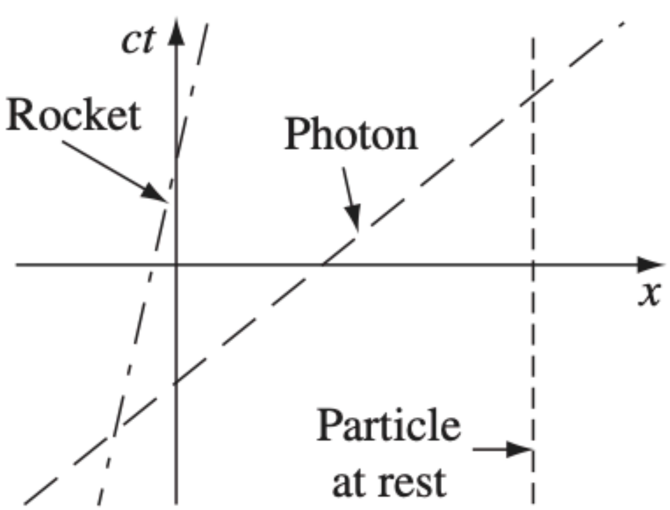

FIGURE 12.21

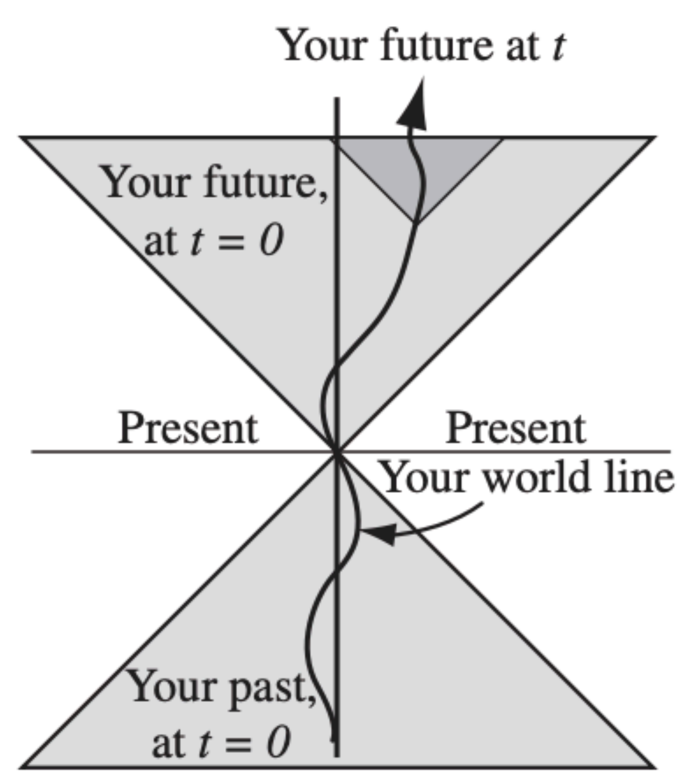

FIGURE 12.22

locus of all points accessible to you. Of course, as time goes on, and you move along your chosen world line, your options progressively narrow: your "future" at any moment is the forward "wedge" constructed at whatever point you find yourself. Meanwhile, the *backward* wedge represents your "past," in the sense that it is the locus of all points from which you might have come. As for the rest (the region outside the forward and backward wedges), this is the generalized "present." You can't *get* there, and you didn't *come* from there. In fact, there's no way you can influence any event in the present (the message would have to travel faster than light); it's a vast expanse of spacetime that is absolutely inaccessible to you.

I've been ignoring the y and z directions. If we include a y axis coming out of the page, the "wedges" become cones—and, with an undrawable z axis, hypercones. Because their boundaries are the trajectories of light rays, we call them the **forward light cone** and the **backward light cone**. Your future, in other words, lies within your forward light cone, your past within your backward light cone.

Notice that the slope of the line connecting two events on a space-time diagram tells you at a glance whether the displacement between them is timelike (slope greater than 1), spacelike (slope less than 1), or lightlike (slope 1). For example, all points in the past and future are timelike with respect to your present location, whereas points in the present are spacelike, and points on the light cone are lightlike.

Hermann Minkowski, who was the first to recognize the full geometrical significance of special relativity, began a famous lecture in 1908 with the words, "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."[12] It's a lovely thought, but you must be careful not to read too much into it. For it is not at all the case that time is "just another coordinate, on the same footing with x, y, and z" (except that for obscure reasons we measure it on clocks instead of rulers). *No:* Time is *utterly different* from the others, and the mark of its distinction is the minus sign in the invariant interval. That minus sign imparts to spacetime a hyperbolic geometry that is much richer than the circular geometry of 3-space.

Under rotations about the z axis, a point P in the xy plane describes a *circle:* the locus of all points a fixed distance $r = \sqrt{x^2 + y^2}$ from the origin (Fig. 12.23). Under Lorentz transformations, however, it is the interval $I = (x^2 - c^2t^2)$ that is preserved, and the locus of all points with a given value of I is a *hyperbola*—or, if we include the y axis, a *hyperboloid of revolution*. When the displacement is *timelike*, it's a "hyperboloid of two sheets" (Fig. 12.24a); when the displacement is *spacelike*, it's a "hyperboloid of one sheet" (Fig. 12.24b). When you perform a Lorentz transformation (that is, when you go into a moving inertial system), the coordinates (x, t) of a given event will change to $(\bar{x}, \bar{t})$, but these new coordinates *will lie on the same hyperbola* as (x, t). By appropriate combinations of Lorentz transformations and rotations, a spot can be moved around at will over the surface of a given hyperboloid, but no amount of transformation will carry it, say, from the upper sheet of the timelike hyperboloid to the lower sheet, or to a spacelike hyperboloid.

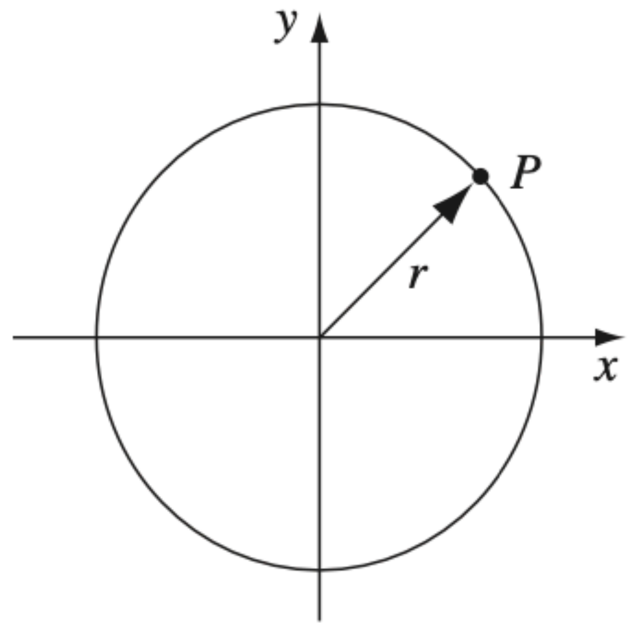

FIGURE 12.23

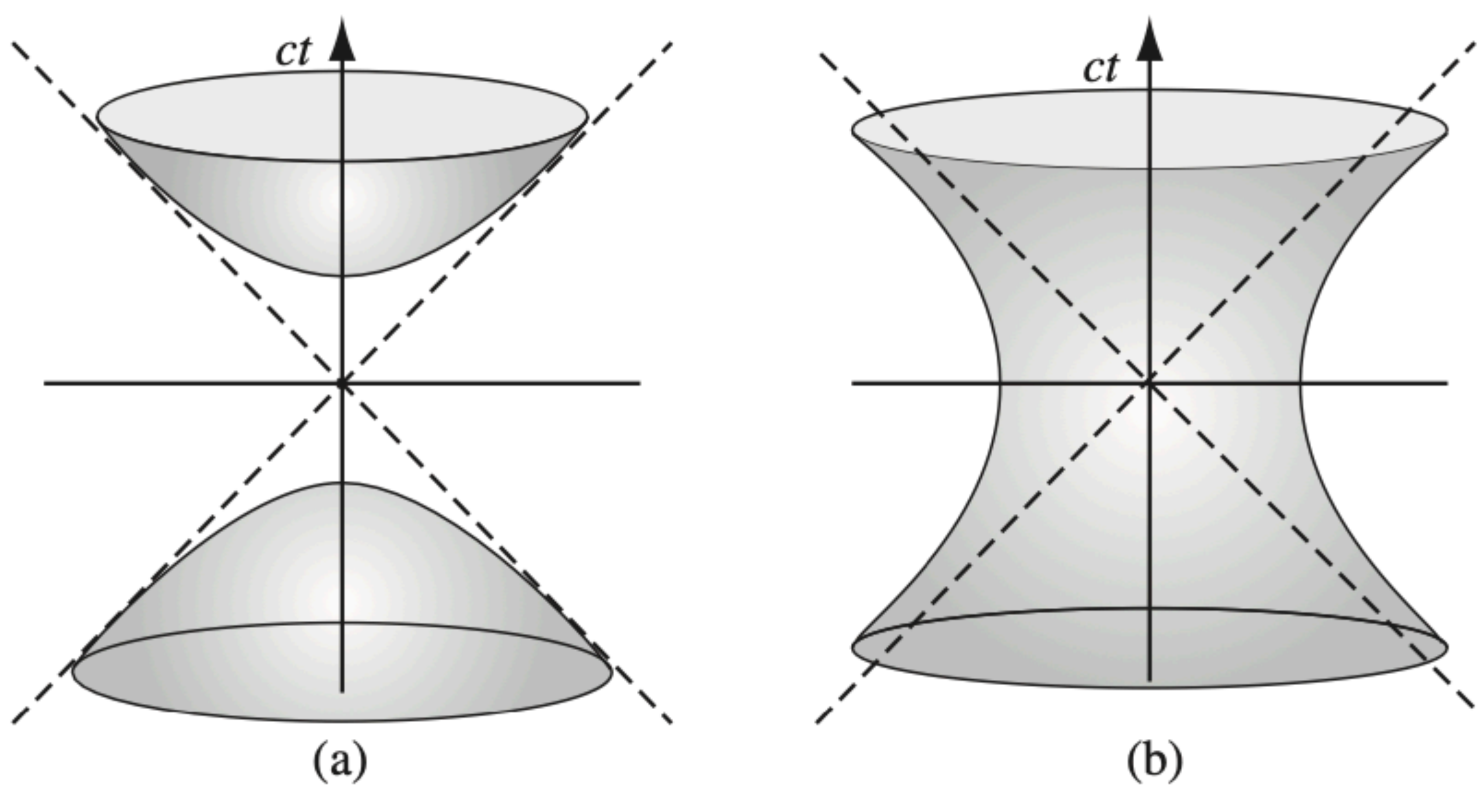

FIGURE 12.24

When we were discussing simultaneity, I showed that the time ordering of two events can, at least in certain cases, be reversed, simply by going into a moving system. But we now see that this is not *always* possible: *If the displacement 4-vector between two events is timelike, their ordering is absolute; if the interval is spacelike, their ordering depends on the inertial system from which they are observed.* In terms of the space-time diagram, an event on the upper sheet of a timelike hyperboloid *definitely* occurred *after* (0, 0), and one on the lower sheet certainly occurred *before;* but an event on a spacelike hyperboloid occurred at positive t, or negative t, depending on your reference frame. This is not an idle curiosity, for it rescues the notion of **causality**, on which all physics is based. If it were *always* possible to reverse the order of two events, then we could never say "A caused B," since a rival observer would retort that B *preceded* A. This embarrassment is avoided, provided the two events are timelike or lightlike separated. And causally related events *are*—otherwise no influence could travel from one to the other. *Conclusion:* The displacement between causally related events is always timelike, and their temporal ordering is the same for all inertial observers.

12.2 ■ RELATIVISTIC MECHANICS

12.2.1 ■ Proper Time and Proper Velocity

As you progress along your world line, your watch runs slow; while the clock on the wall ticks off an interval dt, your watch only advances $d\tau$:

$$d\tau = \sqrt{1 - u^2/c^2}\, dt. \tag{12.37}$$

(I'll use u for the velocity of a particular object—you, in this instance—and reserve v for the relative velocity of two inertial systems.) The time τ your watch registers (or, more generally, the time associated with the moving object) is called **proper time**. (The word suggests a mistranslation of the French *"propre"*, meaning "own.") In some cases, τ may be a more relevant or useful quantity than t. For one thing, proper time is invariant, whereas "ordinary" time depends on the particular reference frame you have in mind.

Now, imagine you're on a flight to Los Angeles, and the pilot announces that the plane's velocity is $\frac{4}{5}c$. What precisely does he mean by "velocity"? Well, of course, he means the displacement divided by the time:

$$\mathbf{u} = \frac{d\mathbf{l}}{dt}, \tag{12.38}$$

and, since he is presumably talking about the velocity relative to ground, both $d\mathbf{l}$ and dt are to be measured by the ground observer. That's the important number to know, if you're concerned about being on time for an appointment in Los Angeles, but if you're wondering whether you'll be hungry on arrival, you might be more interested in the distance covered per unit *proper* time:

$$\boldsymbol{\eta} \equiv \frac{d\mathbf{l}}{d\tau}. \tag{12.39}$$

This hybrid quantity (distance measured on the ground, over time measured in the airplane) is called **proper velocity**; for contrast, I'll call $\mathbf{u}$ the **ordinary velocity**. The two are related by Eq. 12.37:

$$\boldsymbol{\eta} = \frac{1}{\sqrt{1 - u^2/c^2}}\mathbf{u}. \tag{12.40}$$

For speeds much less than c, of course, the difference between ordinary and proper velocity is negligible.

From a theoretical standpoint, however, proper velocity has an enormous advantage over ordinary velocity: it transforms simply, when you go from one inertial system to another. In fact, $\boldsymbol{\eta}$ is the spatial part of a 4-vector,

$$\eta^\mu \equiv \frac{dx^\mu}{d\tau}, \tag{12.41}$$

whose zeroth component is

$$\eta^0 = \frac{dx^0}{d\tau} = c\frac{dt}{d\tau} = \frac{c}{\sqrt{1 - u^2/c^2}}, \tag{12.42}$$

for the numerator, dx^μ, is a displacement 4-vector, while the denominator, $d\tau$, is invariant. Thus, for instance, when you go from system $\mathcal{S}$ to system $\bar{\mathcal{S}}$, moving at speed v along the common $x\bar{x}$ axis,

$$\left.\begin{aligned}\bar{\eta}^0 &= \gamma(\eta^0 - \beta\eta^1),\\ \bar{\eta}^1 &= \gamma(\eta^1 - \beta\eta^0),\\ \bar{\eta}^2 &= \eta^2,\\ \bar{\eta}^3 &= \eta^3.\end{aligned}\right\} \tag{12.43}$$

More generally,

$$\bar{\eta}^\mu = \Lambda^\mu_\nu \eta^\nu; \tag{12.44}$$

η^μ is called the **proper velocity 4-vector**, or simply the **4-velocity**.

By contrast, the transformation rule for *ordinary* velocities is quite cumbersome, as we found in Ex. 12.6 and Prob. 12.14:

$$\left.\begin{aligned}\bar{u}_x &= \frac{d\bar{x}}{d\bar{t}} = \frac{u_x - v}{(1 - vu_x/c^2)},\\ \bar{u}_y &= \frac{d\bar{y}}{d\bar{t}} = \frac{u_y}{\gamma(1 - vu_x/c^2)},\\ \bar{u}_z &= \frac{d\bar{z}}{d\bar{t}} = \frac{u_z}{\gamma(1 - vu_x/c^2)}.\end{aligned}\right\} \tag{12.45}$$

The *reason* for the added complexity is plain: we're obliged to transform both the numerator $d\mathbf{l}$ *and the denominator* dt, whereas for *proper* velocity, the denominator $d\tau$ is invariant, so the ratio inherits the transformation rule of the numerator alone.

12.2.2 ■ Relativistic Energy and Momentum

In classical mechanics, momentum is mass times velocity. I would like to extend this definition to the relativistic domain, but immediately a question arises: Should I use *ordinary* velocity or *proper* velocity? In classical physics, $\boldsymbol{\eta}$ and $\mathbf{u}$ are identical, so there is no *a priori* reason to favor one over the other. However, in the context of relativity it is essential that we use *proper* velocity, for the law of conservation of momentum would be inconsistent with the principle of relativity if we were to define momentum as $m\mathbf{u}$ (see Prob. 12.29). Thus

$$\boxed{\mathbf{p} \equiv m\boldsymbol{\eta} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}};} \tag{12.46}$$

this is the **relativistic momentum** of an object of mass m traveling at (ordinary) velocity $\mathbf{u}$.[13]

Relativistic momentum is the spatial part of a 4-vector,

$$p^\mu \equiv m\eta^\mu, \tag{12.47}$$

and it is natural to ask what the temporal component,

$$p^0 = m\eta^0 = \frac{mc}{\sqrt{1 - u^2/c^2}} \tag{12.48}$$

represents. Einstein identified $p^0 c$ as **relativistic energy**:

$$\boxed{E \equiv \frac{mc^2}{\sqrt{1 - u^2/c^2}};} \tag{12.49}$$

p^μ is called the **energy-momentum 4-vector** (or the **momentum 4-vector**, for short).

Notice that the relativistic energy is nonzero *even when the object is stationary*; we call this **rest energy**:

$$E_{\text{rest}} \equiv mc^2. \tag{12.50}$$

The remainder, which is attributable to the *motion*, is **kinetic energy**

$$E_{\text{kin}} \equiv E - mc^2 = mc^2\left(\frac{1}{\sqrt{1-u^2/c^2}} - 1\right). \tag{12.51}$$

In the nonrelativistic régime ($u \ll c$) the square root can be expanded in powers of u^2/c^2, giving

$$E_{\text{kin}} = \frac{1}{2}mu^2 + \frac{3}{8}\frac{mu^4}{c^2} + \cdots; \tag{12.52}$$

the leading term reproduces the classical formula.

So far, this is all just *notation*. The *physics* resides in the experimental fact that E and $\mathbf{p}$, as defined by Eqs. 12.46 and 12.49, are *conserved:*

> **In every closed[14] system, the total relativistic energy and momentum are conserved.**

Mass is *not* conserved—a fact that has been painfully familiar to everyone since 1945 (though the so-called "conversion of mass into energy" is really a conversion of *rest* energy into *kinetic* energy).

Note the distinction between an **invariant** quantity (same value in all inertial systems) and a **conserved** quantity (same value before and after some process). Mass is invariant but not conserved; energy is conserved but not invariant; electric charge is both conserved *and* invariant; velocity is neither conserved *nor* invariant.

The scalar product of p^μ with itself is

$$p^\mu p_\mu = -(p^0)^2 + (\mathbf{p}\cdot\mathbf{p}) = -m^2c^2, \tag{12.53}$$

as you can quickly check using the result of Prob. 12.26. In terms of the relativistic energy and momentum,

$$\boxed{E^2 - p^2c^2 = m^2c^4.} \tag{12.54}$$

This result is extremely useful, for it enables you to calculate E (if you know $p \equiv |\mathbf{p}|$), or p (knowing E), without ever having to determine the velocity.[15]

12.2.3 ■ Relativistic Kinematics

In this section we'll explore some applications of the conservation laws.

Example 12.7. Two lumps of clay, each of (rest) mass m, collide head-on at $\frac{3}{5}c$ (Fig. 12.26). They stick together. *Question*: what is the mass (M) of the composite lump?

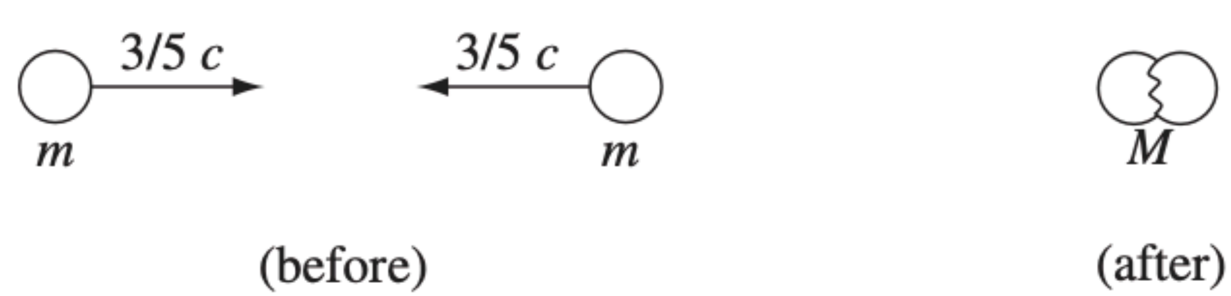

FIGURE 12.26

Solution

In this case conservation of momentum is trivial: zero before, zero after. The energy of each lump prior to the collision is

$$\frac{mc^2}{\sqrt{1-(3/5)^2}} = \frac{5}{4}mc^2,$$

and the energy of the composite lump after the collision is Mc^2 (since it's at rest). So conservation of energy says

$$\frac{5}{4}mc^2 + \frac{5}{4}mc^2 = Mc^2,$$

and hence

$$M = \frac{5}{2}m.$$

Notice that this is *greater* than the sum of the initial masses! Mass was not conserved in this collision; kinetic energy was converted into rest energy, so the mass increased.

In the *classical* analysis of such a collision, we say that kinetic energy was converted into *thermal* energy—the composite lump is *hotter* than the two colliding pieces. This is, of course, true in the relativistic picture too. But what *is* thermal energy? It's the sum total of the random kinetic and potential energies of all the atoms and molecules in the substance. Relativity tells us that these internal energies are represented in the *mass* of the composite object: a hot potato is *heavier* than a cold potato, and a compressed spring is *heavier* than a relaxed spring.

Not by much, it's true—internal energy (U) contributes an amount U/c^2 to the mass, and c^2 is a very large number by everyday standards. You could never get two lumps of clay going anywhere near fast enough to detect the nonconservation of mass in their collision. But in the realm of elementary particles, the effect can be very striking. For example, when the neutral pi meson (mass 2.4×10^{-28} kg) decays into an electron and a positron (each of mass 9.11×10^{-31} kg), the rest energy is converted almost entirely into kinetic energy—less than 1% of the original mass remains.

In classical mechanics, there's no such thing as a massless ($m = 0$) particle—its kinetic energy ($\frac{1}{2}mu^2$) and its momentum ($m\mathbf{u}$) would be zero, you couldn't apply a force to it ($\mathbf{F} = m\mathbf{a}$), and hence (by Newton's third law) it couldn't exert a force on anything else—it's a cipher, as far as physics is concerned. You might at first assume that the same is true in relativity; after all, $\mathbf{p}$ and E are still proportional to m. However, a closer inspection of Eqs. 12.46 and 12.49 reveals a loophole worthy of a congressman: If $u = c$, then the zero in the numerator is balanced by a zero in the denominator, leaving $\mathbf{p}$ and E indeterminate (zero over zero). It is just conceivable, therefore, that a massless particle could carry energy and momentum, *provided it always travels at the speed of light.* Although Eqs. 12.46 and 12.49 would no longer suffice to determine E and $\mathbf{p}$, Eq. 12.54 suggests that the two should be related by

$$E = pc. \tag{12.55}$$

Personally, I would regard this argument as a joke, were it not for the fact that at least one massless particle is known to exist in nature: the photon.[16] Photons *do* travel at the speed of light, and they obey Eq. 12.55.[17] They force us to take the "loophole" seriously. (By the way, you might ask what distinguishes a photon with a lot of energy from one with very little—after all, they have the same mass (zero) and the same speed (c). Relativity offers no answer to this question; curiously, quantum mechanics *does:* According to the Planck formula, $E = h\nu$, where h is **Planck's constant** and ν is the frequency. A blue photon is more energetic than a red one!)

Example 12.8. A pion at rest decays into a muon and a neutrino (Fig. 12.27). Find the energy of the outgoing muon, in terms of the two masses, m_π and m_μ (assume $m_\nu = 0$).

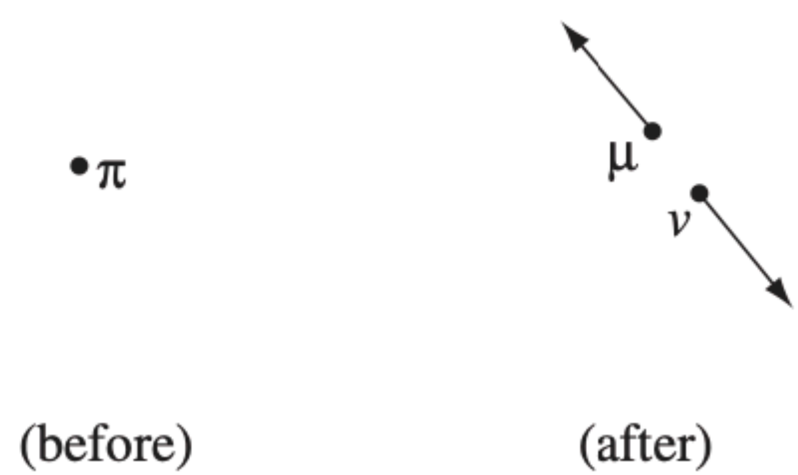

FIGURE 12.27

Solution
In this case,

$$\begin{aligned} E_{\text{before}} &= m_\pi c^2, & \mathbf{p}_{\text{before}} &= 0, \\ E_{\text{after}} &= E_\mu + E_\nu, & \mathbf{p}_{\text{after}} &= \mathbf{p}_\mu + \mathbf{p}_\nu. \end{aligned}$$

Conservation of momentum requires that $\mathbf{p}_\nu = -\mathbf{p}_\mu$. Conservation of energy says that

$$E_\mu + E_\nu = m_\pi c^2.$$

Now, $E_\nu = |\mathbf{p}_\nu|c$, by Eq. 12.55, whereas $|\mathbf{p}_\mu| = \sqrt{E_\mu^2 - m_\mu^2 c^4}\,/c$, by Eq. 12.54, so

$$E_\mu + \sqrt{E_\mu^2 - m_\mu^2 c^4} = m_\pi c^2,$$

from which it follows that

$$E_\mu = \frac{(m_\pi^2 + m_\mu^2)c^2}{2m_\pi}.$$

In a classical collision, momentum and mass are always conserved, whereas kinetic energy, in general, is not. A "sticky" collision generates heat at the expense of kinetic energy; an "explosive" collision generates kinetic energy at the expense of chemical energy (or some other kind). If the kinetic energy *is* conserved, as in the ideal collision of the two billiard balls, we call the process "elastic." In the relativistic case, momentum and total energy are always conserved, but mass and kinetic energy, in general, are not. Once again, we call the process **elastic** if kinetic energy is conserved. In such a case the rest energy (being the total minus the kinetic) is *also* conserved, and therefore so too is the mass. In practice, this means that the *same particles* come out as went in. Examples 12.7 and 12.8 were inelastic processes; the next one is elastic.

Example 12.9. Compton scattering. A photon of energy E_0 "bounces" off an electron, initially at rest. Find the energy E of the outgoing photon, as a function of the **scattering angle** θ (see Fig. 12.28).

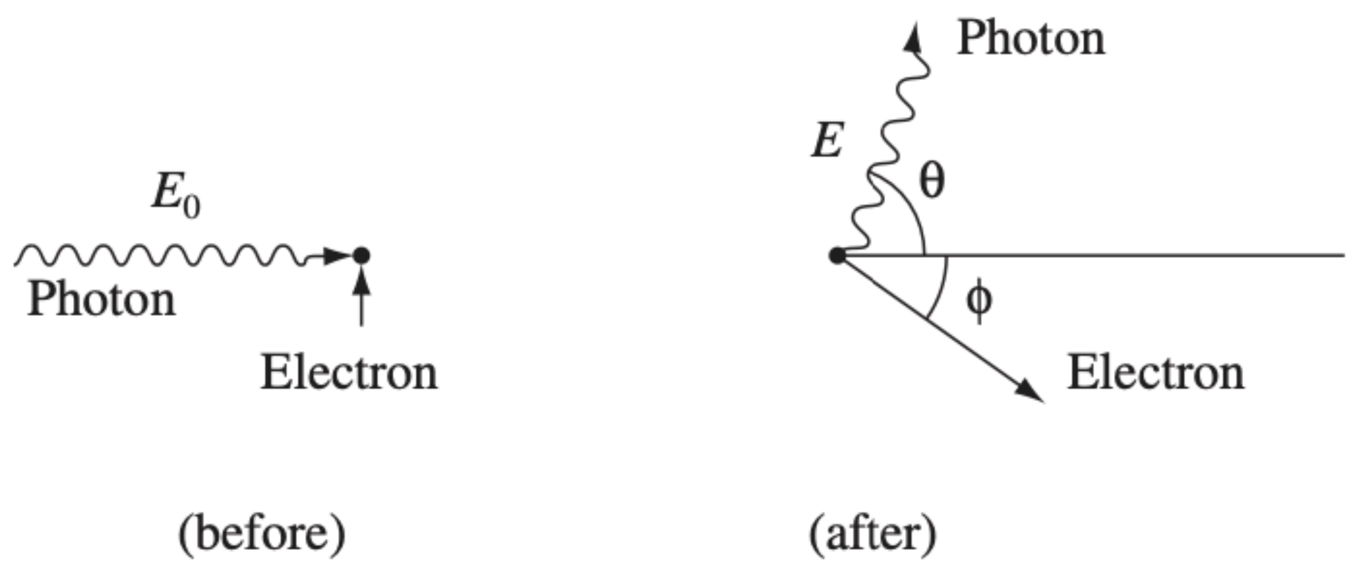

FIGURE 12.28

Solution

Conservation of momentum in the "vertical" direction gives $p_e \sin\phi = p_p \sin\theta$, or, since $p_p = E/c$,

$$\sin\phi = \frac{E}{p_e c}\sin\theta.$$

Conservation of momentum in the "horizontal" direction gives

$$\frac{E_0}{c} = p_p\cos\theta + p_e\cos\phi = \frac{E}{c}\cos\theta + p_e\sqrt{1 - \left(\frac{E}{p_e c}\sin\theta\right)^2},$$

or

$$p_e^2 c^2 = (E_0 - E\cos\theta)^2 + E^2\sin^2\theta = E_0^2 - 2E_0 E\cos\theta + E^2.$$

Finally, conservation of energy says that

$$\begin{aligned} E_0 + mc^2 &= E + E_e = E + \sqrt{m^2c^4 + p_e^2c^2} \\ &= E + \sqrt{m^2c^4 + E_0^2 - 2E_0E\cos\theta + E^2}. \end{aligned}$$

Solving for E, I find that

$$E = \frac{1}{(1-\cos\theta)/mc^2 + (1/E_0)}. \tag{12.56}$$

The answer looks nicer when expressed in terms of photon wavelength:

$$E = h\nu = \frac{hc}{\lambda},$$

so

$$\lambda = \lambda_0 + \frac{h}{mc}(1-\cos\theta). \tag{12.57}$$

The quantity (h/mc) is called the **Compton wavelength** of the electron.

12.2.4 ■ Relativistic Dynamics

Newton's first law is built into the principle of relativity. His second law, in the form

$$\boxed{\mathbf{F} = \frac{d\mathbf{p}}{dt},} \tag{12.59}$$

retains its validity in relativistic mechanics, *provided we use the relativistic momentum.*

Example 12.10. Motion under a constant force. A particle of mass m is subject to a constant force F. If it starts from rest at the origin, at time $t = 0$, find its position (x), as a function of time.

Solution

$$\frac{dp}{dt} = F \quad \Rightarrow \quad p = Ft + \text{constant},$$

but since $p = 0$ at $t = 0$, the constant must be zero, and hence

$$p = \frac{mu}{\sqrt{1 - u^2/c^2}} = Ft.$$

Solving for u, we obtain

$$u = \frac{(F/m)t}{\sqrt{1 + (Ft/mc)^2}}. \tag{12.60}$$

To complete the problem we must integrate again:

$$x(t) = \frac{F}{m}\int_0^t \frac{t'}{\sqrt{1+(Ft'/mc)^2}}\,dt'$$

$$= \frac{mc^2}{F}\sqrt{1+(Ft'/mc)^2}\bigg|_0^t = \frac{mc^2}{F}\left[\sqrt{1+(Ft/mc)^2}-1\right]. \quad (12.61)$$

In place of the classical parabola, $x(t) = (F/2m)t^2$, the graph is a *hyperbola* (Fig. 12.30); for this reason, motion under a constant force is often called **hyperbolic motion.** It occurs, for example, when a charged particle is placed in a uniform electric field.

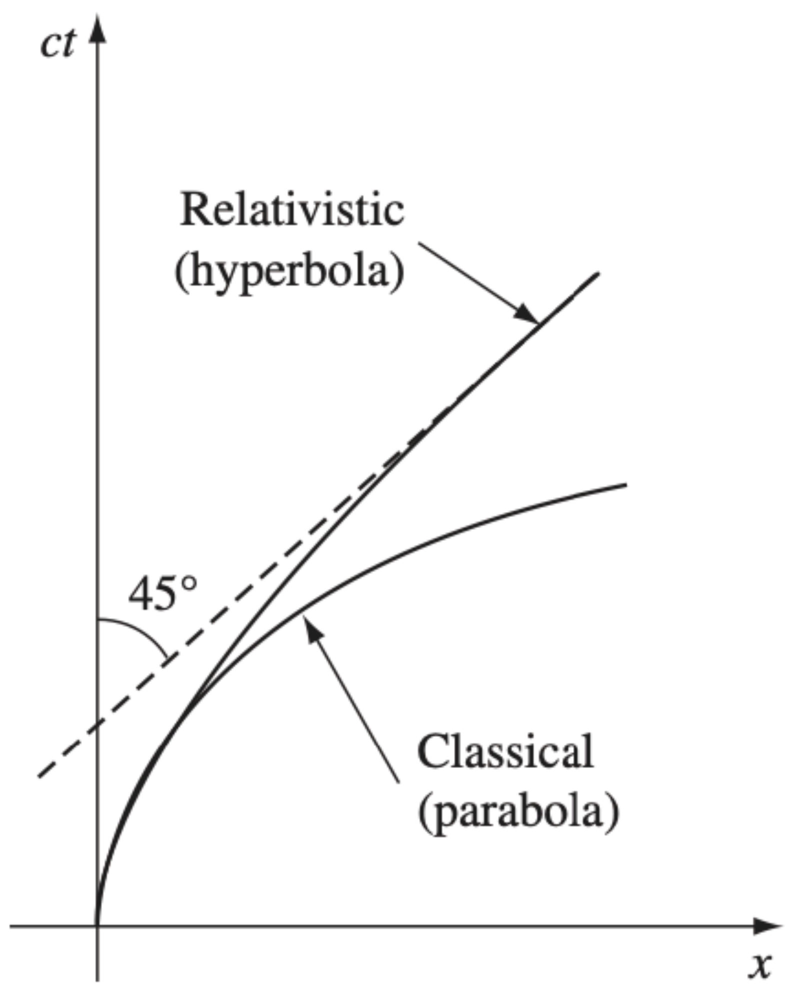

FIGURE 12.30

Work, as always, is the line integral of the force:

$$W \equiv \int \mathbf{F} \cdot d\mathbf{l}. \tag{12.62}$$

The **work-energy theorem** ("the net work done on a particle equals the increase in its kinetic energy") holds relativistically:

$$W = \int \frac{d\mathbf{p}}{dt} \cdot d\mathbf{l} = \int \frac{d\mathbf{p}}{dt} \cdot \frac{d\mathbf{l}}{dt}\, dt = \int \frac{d\mathbf{p}}{dt} \cdot \mathbf{u}\, dt,$$

while

$$\begin{aligned} \frac{d\mathbf{p}}{dt} \cdot \mathbf{u} &= \frac{d}{dt}\left(\frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}}\right) \cdot \mathbf{u} \\ &= \frac{m\mathbf{u}}{(1 - u^2/c^2)^{3/2}} \cdot \frac{d\mathbf{u}}{dt} = \frac{d}{dt}\left(\frac{mc^2}{\sqrt{1 - u^2/c^2}}\right) = \frac{dE}{dt}, \end{aligned} \tag{12.63}$$

so

$$W = \int \frac{dE}{dt}\, dt = E_{\text{final}} - E_{\text{initial}}. \tag{12.64}$$

(Since the rest energy is constant, it doesn't matter whether we use the total energy, here, or the kinetic energy.)

Unlike the first two, Newton's *third* law does not, in general, extend to the relativistic domain. Indeed, if the two objects in question are separated in space, the third law is incompatible with the relativity of simultaneity. For suppose the force of A on B at some instant t is $\mathbf{F}(t)$, and the force of B on A at the same instant is $-\mathbf{F}(t)$; then the third law applies *in this reference frame.* But a moving observer will report that these equal and opposite forces occurred at *different times;* in his system, therefore, the third law is violated. Only in the case of contact interactions, where the two forces are applied at the same physical point (and in the trivial case where the forces are constant) can the third law be retained.

Because **F** is the derivative of momentum with respect to *ordinary* time, it shares the ugly behavior of (ordinary) velocity, when you go from one inertial system to another: both the numerator *and the denominator* must be transformed. Thus,[18]

$$\bar{F}_y = \frac{d\bar{p}_y}{d\bar{t}} = \frac{dp_y}{\gamma\, dt - \dfrac{\gamma\beta}{c}\, dx} = \frac{dp_y/dt}{\gamma\left(1 - \dfrac{\beta}{c}\dfrac{dx}{dt}\right)} = \frac{F_y}{\gamma(1 - \beta u_x/c)}, \tag{12.65}$$

and similarly for the z component:

$$\bar{F}_z = \frac{F_z}{\gamma(1 - \beta u_x/c)}.$$

The x component is even worse:

$$\bar{F}_x = \frac{d\bar{p}_x}{d\bar{t}} = \frac{\gamma\, dp_x - \gamma\beta\, dp^0}{\gamma\, dt - \dfrac{\gamma\beta}{c}\, dx} = \frac{\dfrac{dp_x}{dt} - \beta\dfrac{dp^0}{dt}}{1 - \dfrac{\beta}{c}\dfrac{dx}{dt}} = \frac{F_x - \dfrac{\beta}{c}\left(\dfrac{dE}{dt}\right)}{1 - \beta u_x/c}.$$

We calculated dE/dt in Eq. 12.63; putting that in,

$$\bar{F}_x = \frac{F_x - \beta(\mathbf{u}\cdot\mathbf{F})/c}{1 - \beta u_x/c}. \tag{12.66}$$

In one special case these equations are reasonably tractable: *If the particle is* (instantaneously) *at rest in* $\mathcal{S}$, so that $\mathbf{u} = 0$, then

$$\bar{\mathbf{F}}_\perp = \frac{1}{\gamma}\mathbf{F}_\perp, \quad \bar{F}_\parallel = F_\parallel. \tag{12.67}$$

That is, the component of **F** *parallel* to the motion of $\bar{S}$ is unchanged, whereas perpendicular components are divided by γ.

It has perhaps occurred to you that we could avoid the bad transformation behavior of **F** by introducing a "proper" force, analogous to proper velocity, which would be the derivative of momentum with respect to *proper* time:

$$K^\mu \equiv \frac{dp^\mu}{d\tau}. \tag{12.68}$$

This is called the **Minkowski force**; it is plainly a 4-vector, since p^μ is a 4-vector and proper time is invariant. The spatial components of K^μ are related to the "ordinary" force by

$$\mathbf{K} = \left(\frac{dt}{d\tau}\right)\frac{d\mathbf{p}}{dt} = \frac{1}{\sqrt{1-u^2/c^2}}\mathbf{F}, \tag{12.69}$$

while the zeroth component,

$$K^0 = \frac{dp^0}{d\tau} = \frac{1}{c}\frac{dE}{d\tau}, \tag{12.70}$$

is, apart from the $1/c$, the (proper) rate at which the energy of the particle increases—in other words, the (proper) *power* delivered to the particle.

Relativistic dynamics can be formulated in terms of the ordinary force *or* in terms of the Minkowski force. The latter is generally much neater, but since in the long run we are interested in the particle's trajectory as a function of *ordinary* time, the former is often more useful. When we wish to generalize some classical force law, such as Lorentz's, to the relativistic domain, the question arises: Does the classical formula correspond to the ordinary force or to the Minkowski force? In other words, should we write

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

or should it rather be

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

Since proper time and ordinary time are identical in classical physics, there is no way at this stage to decide the issue. The Lorentz force, as it turns out, is an *ordinary* force—later on I'll explain why this is so, and show you how to construct the electromagnetic Minkowski force.

Example 12.11. The typical trajectory of a charged particle in a uniform magnetic field is **cyclotron motion** (Fig. 12.31). The magnetic force pointing toward the center,

$$F = QuB,$$

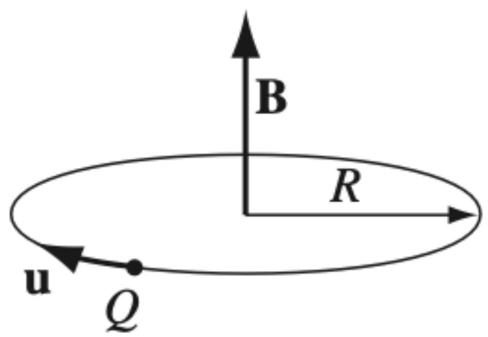

FIGURE 12.31

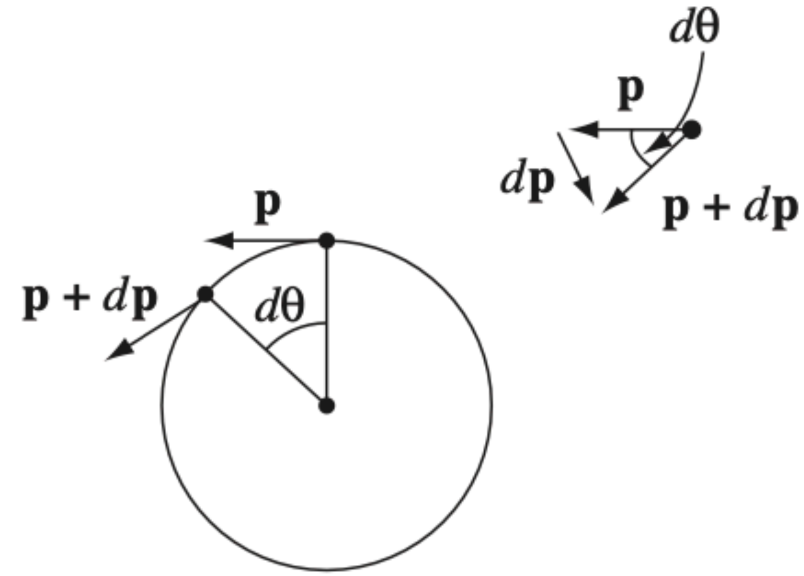

FIGURE 12.32

provides the centripetal acceleration necessary to sustain circular motion. Beware, however—in special relativity the centripetal force is *not* mu^2/R, as in classical mechanics. Rather, as you can see from Fig. 12.32, $dp = p\,d\theta$, so

$$F = \frac{dp}{dt} = p\frac{d\theta}{dt} = p\frac{u}{R}.$$

(Classically, of course, $p = mu$, so $F = mu^2/R$.) Thus,

$$QuB = p\frac{u}{R},$$

or

$$p = QBR. \tag{12.71}$$

In this form, the relativistic cyclotron formula is identical to the nonrelativistic one, Eq. 5.3—the only difference is that p is now the *relativistic* momentum.

In classical mechanics, the total momentum (**P**) of a collection of interacting particles can be expressed as the total mass (M) times the velocity of the center-of-mass:

$$\mathbf{P} = M\frac{d\mathbf{R}_m}{dt}.$$

In relativity the center-of-mass ($\mathbf{R}_m = \frac{1}{M}\sum m_i\mathbf{r}_i$) is replaced by the **center-of-energy** ($\mathbf{R}_e = \frac{1}{E}\sum E_i\mathbf{r}_i$, where E is the total energy), and M by E/c^2:

$$\mathbf{P} = \frac{E}{c^2}\frac{d\mathbf{R}_e}{dt}. \tag{12.72}$$

P now includes *all* forms of momentum, and E all forms of energy—not just mechanical, but also whatever may be stored in the fields.[19]

Example 12.12. In Example 8.3 we found that the momentum stored in the fields of a coaxial cable is not zero, even though the cable itself is at rest. At the time, this seemed paradoxical. However, energy is being transported, from the battery to the resistor, and hence the center-of-energy is in motion. Indeed, if the battery is at $z = 0$, so the resistor is at $z = l$, then $\mathbf{R}_e = (E_0\mathbf{R}_0 + E_R l\,\hat{\mathbf{z}})/E$, where E_R is the energy in the resistor, E_0 is the rest of the energy, and $\mathbf{R}_0$ is the center-of-energy of E_0, so

$$\frac{d\mathbf{R}_e}{dt} = \frac{(dE_R/dt)l}{E}\,\hat{\mathbf{z}} = \frac{IVl}{E}\,\hat{\mathbf{z}}.$$

According to Eq. 12.72, then, the total momentum should be

$$\mathbf{P} = \frac{IVl}{c^2}\,\hat{\mathbf{z}},$$

which is exactly the momentum in the fields, as calculated in Example 8.3.

If this still seems strange to you, imagine a shoe-box, with a marble inside that we cannot see. The box is at rest, but the marble is rolling from one end to the other. Is there momentum in this system? Yes, of course, even though the box is stationary—there is the momentum of the marble. In the case of the coaxial cable, no actual *object* is in motion (well, the electrons are, but there are just as many of them going one way as the other, and their net momentum is zero), but energy is flowing from one end to the other, and in relativity *all* forms of energy in motion, not just rest energy (mass), constitute momentum. The "marble" (in this analogy) is the electromagnetic field, which transports energy, and therefore contributes momentum ... *even though the fields themselves are perfectly static!* [20]

In the following example, the center of energy is at rest, so the total momentum must be zero (Eq. 12.72). But the (static) electromagnetic fields do carry momentum, and the problem is to locate the compensating mechanical momentum.

Example 12.13. As a model for a magnetic dipole **m**, consider a rectangular loop of wire carrying a steady current I. Picture the current as a stream of non-interacting positive charges that move freely within the wire. When a uniform electric field **E** is applied (Fig. 12.33), the charges accelerate in the left segment and decelerate in the right one.[21] Find the total momentum of all the charges in the loop.

Solution

The momenta of the left and right segments cancel, so we need only consider the top and the bottom. Say there are N_+ charges in the top segment, going at speed

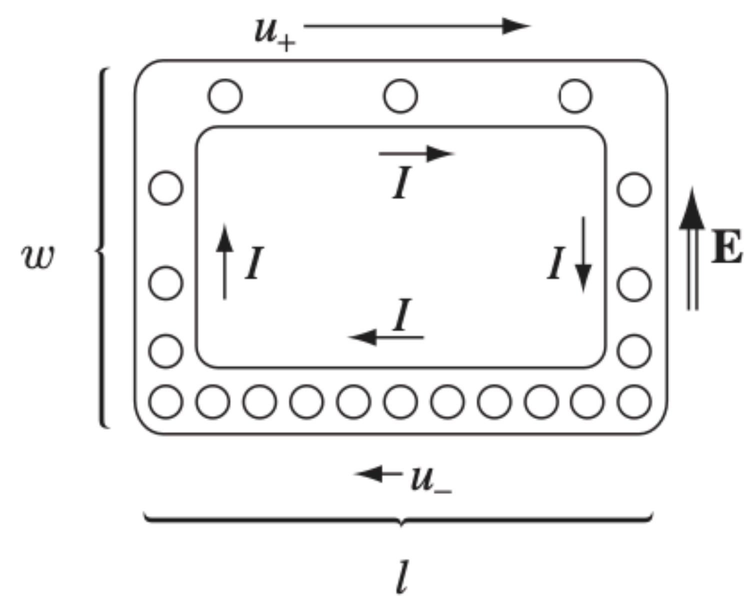

FIGURE 12.33

u_+ to the right, and N_- charges in the lower segment, going at (slower) speed u_- to the left. The current ($I = \lambda u$) is the same in all four segments (or else charge would be piling up somewhere); in particular,

$$I = \frac{QN_+}{l} u_+ = \frac{QN_-}{l} u_-, \quad \text{so} \quad N_\pm u_\pm = \frac{Il}{Q},$$

where Q is the charge of each particle, and l is the length of the rectangle. *Classically,* the momentum of a single particle is $\mathbf{p} = M\mathbf{u}$ (where M is its mass), and the total momentum (to the right) is

$$p_{\text{classical}} = MN_+u_+ - MN_-u_- = M\frac{Il}{Q} - M\frac{Il}{Q} = 0,$$

as one would certainly expect (after all, the loop as a whole is not moving). But *relativistically*, $\mathbf{p} = \gamma M\mathbf{u}$, and we get

$$p = \gamma_+ MN_+u_+ - \gamma_- MN_-u_- = \frac{MIl}{Q}(\gamma_+ - \gamma_-),$$

which is *not* zero, because the particles in the upper segment are moving faster.

In fact, the gain in energy (γMc^2), as a particle goes up the left segment, is equal to the work done by the electric force, QEw, where w is the height of the rectangle, so

$$\gamma_+ - \gamma_- = \frac{QEw}{Mc^2},$$

and hence

$$p = \frac{IlEw}{c^2}.$$

But Ilw is the magnetic dipole moment of the loop; as vectors, $\mathbf{m}$ points into the page and $\mathbf{p}$ is to the right, so

$$\mathbf{p} = \frac{1}{c^2}(\mathbf{m} \times \mathbf{E}). \tag{12.73}$$

Thus a magnetic dipole at rest in an electric field carries linear momentum, *even though it is not moving!* This so-called **hidden momentum** is strictly relativistic, and purely mechanical; it precisely cancels the electromagnetic momentum stored in the fields (Eq. 8.45).[22]

12.3 ■ RELATIVISTIC ELECTRODYNAMICS

12.3.1 ■ Magnetism as a Relativistic Phenomenon

Unlike Newtonian mechanics, classical electrodynamics is *already* consistent with special relativity. Maxwell's equations and the Lorentz force law can be applied legitimately in any inertial system. Of course, what one observer interprets as an electrical process another may regard as magnetic, but the actual particle motions they predict will be identical. To the extent that this did *not* work out for Lorentz and others, who studied the question in the late nineteenth century, the fault lay with the nonrelativistic mechanics they used, not with the electrodynamics. Having corrected Newtonian mechanics, we are now in a position to develop a complete and consistent formulation of relativistic electrodynamics. I emphasize that we will not be changing the rules of electrodynamics in the slightest—rather, we will be *expressing* these rules in a notation that exposes and illuminates their relativistic character. As we go along, I shall pause now and then to rederive, using the Lorentz transformations, results obtained earlier by more laborious means. But the main purpose of this section is to provide you with a deeper understanding of the structure of electrodynamics—laws that had seemed arbitrary and unrelated before take on a kind of coherence and inevitability when approached from the point of view of relativity.

To begin with, I'd like to show you why there *had* to be such a thing as magnetism, given electrostatics and relativity, and how, in particular, you can calculate the magnetic force between a current-carrying wire and a moving charge without ever invoking the laws of magnetism.[23] Suppose you had a string of positive charges moving along to the right at speed v. I'll assume the charges are close enough together so that we may treat them as a continuous line charge λ. Superimposed on this positive string is a negative one, $-\lambda$ proceeding to the left at the same speed v. We have, then, a net current to the right, of magnitude

$$I = 2\lambda v. \tag{12.76}$$

Meanwhile, a distance s away there is a point charge q traveling to the right at speed $u < v$ (Fig. 12.34a). Because the two line charges cancel, there is *no electrical force on* q in this system ($\mathcal{S}$).

However, let's examine the same situation from the point of view of system $\bar{\mathcal{S}}$, which moves to the right with speed u (Fig. 12.34b). In this reference frame, q is at rest. By the Einstein velocity addition rule, the velocities of the positive and negative lines are now

$$v_{\pm} = \frac{v \mp u}{1 \mp vu/c^2}. \tag{12.77}$$

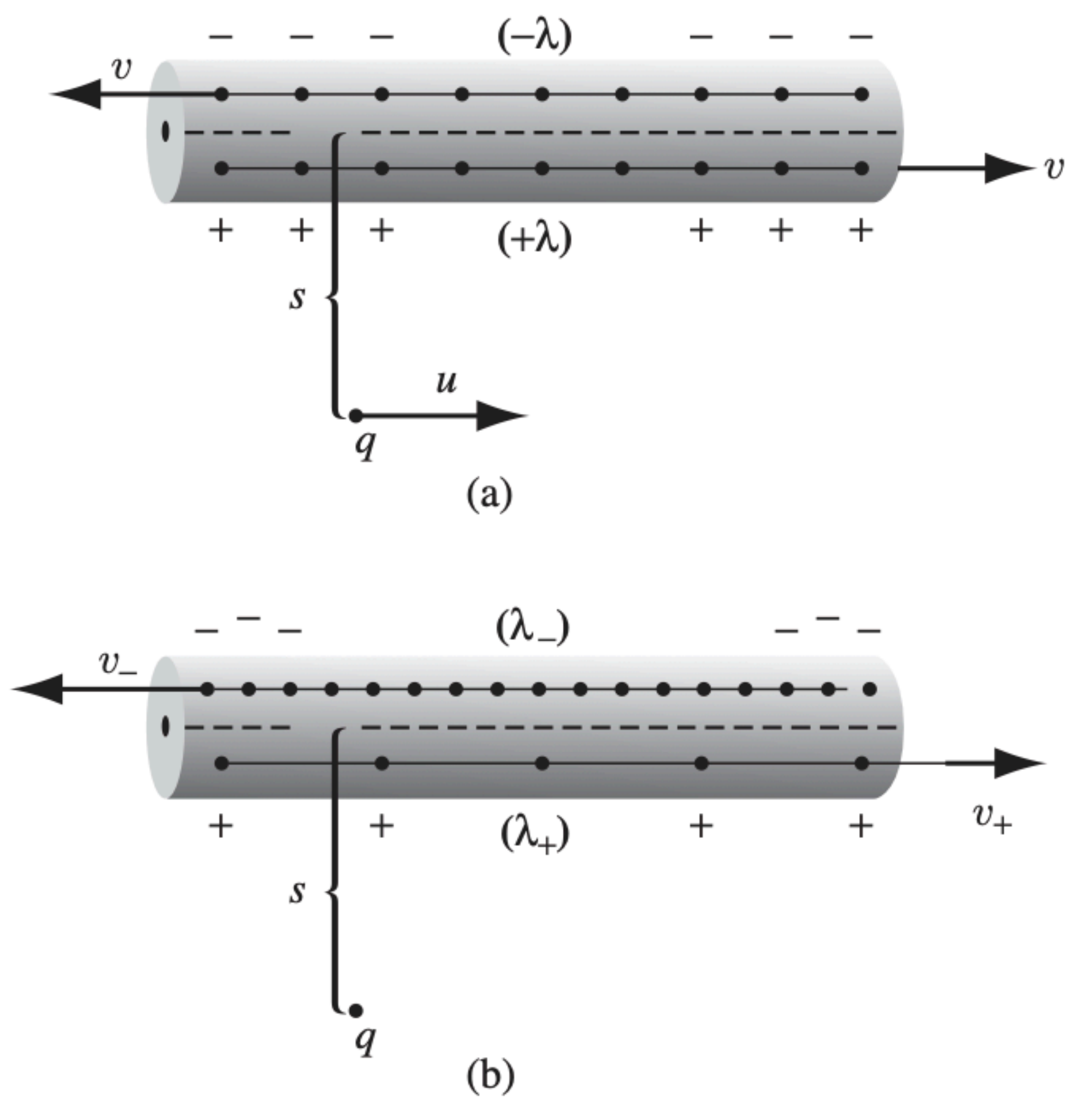

FIGURE 12.34

Because v_- is greater than v_+, the Lorentz contraction of the spacing between negative charges is more severe than that between positive charges; *in this frame,* therefore, *the wire carries a net negative charge!* In fact,

$$\lambda_\pm = \pm(\gamma_\pm)\lambda_0, \tag{12.78}$$

where

$$\gamma_\pm = \frac{1}{\sqrt{1 - v_\pm^2/c^2}}, \tag{12.79}$$

and λ_0 is the charge density of the positive line in its own rest system. That's not the same as λ, of course—in $\mathcal{S}$ they're already moving at speed v, so

$$\lambda = \gamma\lambda_0, \tag{12.80}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}. \tag{12.81}$$

It takes some algebra to put $\gamma_\pm$ into simple form:

$$\begin{aligned}\gamma_\pm &= \frac{1}{\sqrt{1 - \frac{1}{c^2}(v \mp u)^2(1 \mp vu/c^2)^{-2}}} = \frac{c^2 \mp uv}{\sqrt{(c^2 \mp uv)^2 - c^2(v \mp u)^2}} \\ &= \frac{c^2 \mp uv}{\sqrt{(c^2 - v^2)(c^2 - u^2)}} = \gamma\frac{1 \mp uv/c^2}{\sqrt{1 - u^2/c^2}}.\end{aligned} \tag{12.82}$$

The net line charge in $\bar{S}$, then, is

$$\lambda_{\text{tot}} = \lambda_+ + \lambda_- = \lambda_0(\gamma_+ - \gamma_-) = \frac{-2\lambda u v}{c^2\sqrt{1 - u^2/c^2}}. \tag{12.83}$$

Conclusion: As a result of unequal Lorentz contraction of the positive and negative lines, a current-carrying wire that is electrically neutral in one inertial system will be charged in another.

Now, a line charge λ_{tot} sets up an *electric* field

$$E = \frac{\lambda_{\text{tot}}}{2\pi\epsilon_0 s},$$

so there is an electrical force on q in $\bar{S}$, to wit:

$$\bar{F} = qE = -\frac{\lambda v}{\pi\epsilon_0 c^2 s} \frac{qu}{\sqrt{1 - u^2/c^2}}. \tag{12.84}$$

But if there's a force on q in $\bar{S}$, there must be one in S; in fact, we can calculate it by using the transformation rules for forces. Since q is at rest in $\bar{S}$, and $\bar{F}$ is perpendicular to u, the force in S is given by Eq. 12.67:

$$F = \sqrt{1 - u^2/c^2}\,\bar{F} = -\frac{\lambda v}{\pi\epsilon_0 c^2}\frac{qu}{s}. \tag{12.85}$$

The charge is attracted toward the wire by a force that is purely electrical in $\bar{\mathcal{S}}$ (where the wire is charged, and q is at rest), but distinctly *non*electrical in $\mathcal{S}$ (where the wire is neutral). Taken together, then, electrostatics and relativity imply the existence of another force. This "other force" is, of course, *magnetic.* In fact, we can cast Eq. 12.85 into more familiar form by using $c^2 = (\epsilon_0\mu_0)^{-1}$ and expressing λv in terms of the current (Eq. 12.76):

$$F = -qu\left(\frac{\mu_0 I}{2\pi s}\right). \tag{12.86}$$

The term in parentheses is the magnetic field of a long straight wire, and the force is precisely what we would have obtained by using the Lorentz force law in system $\mathcal{S}$.

12.3.2 ■ How the Fields Transform

We have learned, in various special cases, that one observer's electric field is another's magnetic field. It would be nice to know the *general* transformation rules for electromagnetic fields: Given the fields in $\mathcal{S}$, what are the fields in $\bar{\mathcal{S}}$? Your first guess might be that $\mathbf{E}$ is the spatial part of one 4-vector and $\mathbf{B}$ the spatial part of another. But your guess would be wrong—it's more complicated than that. Let me begin by making explicit an assumption that was already used implicitly in Sect. 12.3.1: *Charge is invariant.* Like mass, but unlike energy, the charge of a particle is a fixed number, independent of how fast it happens to be moving. We shall assume also that the transformation rules are the same no matter how the fields were produced—electric fields associated with changing magnetic fields transform the same way as those set up by stationary charges. Were this not the case we'd have to abandon the field formulation altogether, for it is the essence of a field theory that the fields at a given point tell you *all there is to know,* electromagnetically, about that point; you do *not* have to append extra information regarding their source.

With this in mind, consider the simplest possible electric field: the uniform field in the region between the plates of a large parallel-plate capacitor (Fig. 12.35a). Say the capacitor is at rest in $\mathcal{S}_0$ and carries surface charges $\pm\sigma_0$. Then

$$\mathbf{E}_0 = \frac{\sigma_0}{\epsilon_0}\,\hat{\mathbf{y}}. \tag{12.87}$$

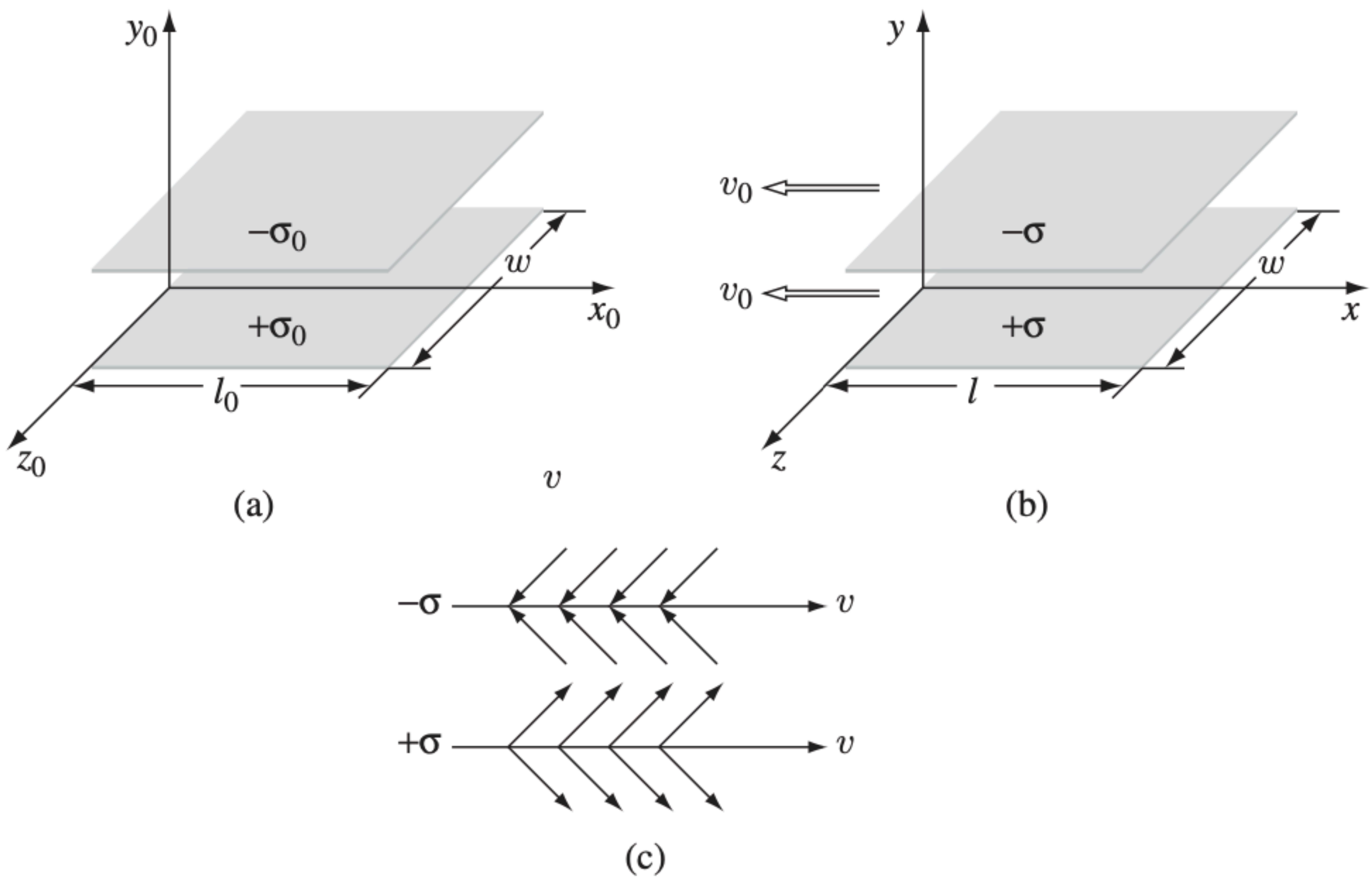

FIGURE 12.35

What if we examine this same capacitor from system $\mathcal{S}$, moving to the right at speed v_0 (Fig. 12.35b)? In this system the plates are moving to the left, but the field still takes the form

$$\mathbf{E} = \frac{\sigma}{\epsilon_0}\,\hat{\mathbf{y}}; \tag{12.88}$$

the only difference is the value of the surface charge σ. [Wait a minute! *Is* that the only difference? The formula $E = \sigma/\epsilon_0$ for a parallel plate capacitor came from Gauss's law, and whereas Gauss's law is perfectly valid for moving charges, this application also relies on symmetry. Are we sure that the field is still perpendicular to the plates? What if the field of a moving plane *tilts,* say, along the direction of motion, as in Fig. 12.35c? Well, *even if it did* (it doesn't), the field between the plates, being the superposition of the $+\sigma$ field and the $-\sigma$ field, would still run perpendicular to the plates (changing the sign of the charge reverses the direction of the field, and the vector sum kills off the parallel components).]

Now, the total charge on each plate is invariant, and the *width* (w) is unchanged, but the *length* (l) is Lorentz-contracted by a factor of

$$\gamma_0 = \frac{1}{\sqrt{1 - v_0^2/c^2}}, \tag{12.89}$$

so the charge per unit area is *increased* by a factor of γ_0:

$$\sigma = \gamma_0 \sigma_0. \tag{12.90}$$

Accordingly,

$$\mathbf{E}^{\perp} = \gamma_0 \mathbf{E}_0{}^{\perp}. \tag{12.91}$$

I have put in the superscript $\perp$ to make it clear that this rule pertains to components of $\mathbf{E}$ that are *perpendicular* to the direction of motion of $\mathcal{S}$. To get the rule for *parallel* components, consider the capacitor lined up with the yz plane

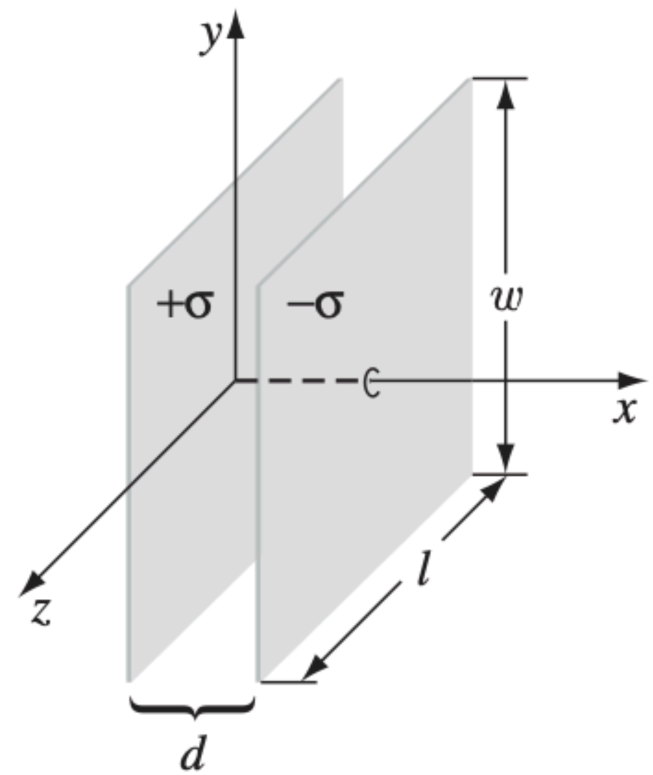

FIGURE 12.36

(Fig. 12.36). This time it is the plate separation (d) that is Lorentz-contracted, whereas l and w (and hence also σ) are the same in both frames. Since the field does not depend on d, it follows that

$$E^{\parallel} = E_0{}^{\parallel}. \tag{12.92}$$

Example 12.14. Electric field of a point charge in uniform motion. A point charge q is at rest at the origin in system $\mathcal{S}_0$. *Question*: What is the electric field of this same charge in system $\mathcal{S}$, which moves to the right at speed v_0 relative to $\mathcal{S}_0$?

Solution

In $\mathcal{S}_0$, the field is

$$\mathbf{E}_0 = \frac{1}{4\pi\epsilon_0}\frac{q}{r_0^2}\,\hat{\mathbf{r}}_0,$$

or

$$\begin{cases} E_{x0} &= \dfrac{1}{4\pi\epsilon_0}\dfrac{qx_0}{(x_0^2+y_0^2+z_0^2)^{3/2}}, \\[2ex] E_{y0} &= \dfrac{1}{4\pi\epsilon_0}\dfrac{qy_0}{(x_0^2+y_0^2+z_0^2)^{3/2}}, \\[2ex] E_{z0} &= \dfrac{1}{4\pi\epsilon_0}\dfrac{qz_0}{(x_0^2+y_0^2+z_0^2)^{3/2}}. \end{cases}$$

From the transformation rules (Eqs. 12.91 and 12.92), we have

$$\begin{cases} E_x = E_{x0} &= \dfrac{1}{4\pi\epsilon_0}\dfrac{qx_0}{(x_0^2+y_0^2+z_0^2)^{3/2}}, \\[2ex] E_y = \gamma_0 E_{y0} &= \dfrac{1}{4\pi\epsilon_0}\dfrac{\gamma_0 q y_0}{(x_0^2+y_0^2+z_0^2)^{3/2}}, \\[2ex] E_z = \gamma_0 E_{z0} &= \dfrac{1}{4\pi\epsilon_0}\dfrac{\gamma_0 q z_0}{(x_0^2+y_0^2+z_0^2)^{3/2}}. \end{cases}$$

These are still expressed in terms of the $\mathcal{S}_0$ coordinates (x_0, y_0, z_0) of the field point (P); I'd prefer to write them in terms of the $\mathcal{S}$ coordinates of P. From the Lorentz transformations (or, actually, the inverse transformations),

$$\begin{cases} x_0 = \gamma_0(x + v_0 t) = \gamma_0 R_x, \\ y_0 = y = R_y, \\ z_0 = z = R_z, \end{cases}$$

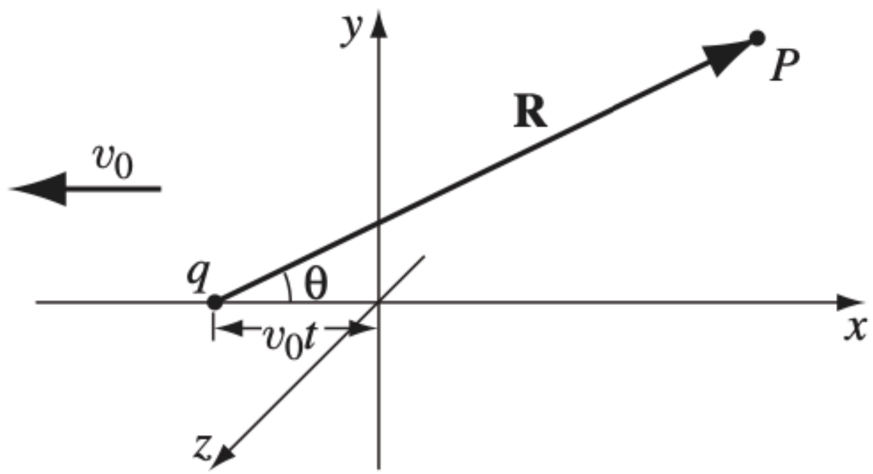

FIGURE 12.37

where **R** is the vector from q to P (Fig. 12.37). Thus

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{\gamma_0 q \mathbf{R}}{(\gamma_0^2 R^2 \cos^2\theta + R^2 \sin^2\theta)^{3/2}}$$

$$= \frac{1}{4\pi\epsilon_0} \frac{q(1 - v_0^2/c^2)}{[1 - (v_0^2/c^2)\sin^2\theta]^{3/2}} \frac{\hat{\mathbf{R}}}{R^2}. \tag{12.93}$$

This, then, is the field of a charge in uniform motion; we got the same result in Chapter 10 using the retarded potentials (Eq. 10.75). The present derivation is far more efficient, and sheds some light on the remarkable fact that the field points away from the instantaneous (as opposed to the retarded) position of the charge: E_x gets a factor of γ_0 from the Lorentz transformation of the *coordinates;* E_y and E_z pick up theirs from the transformation of the *field.* It's the balancing of these two γ_0's that leaves **E** parallel to **R**.

But Eqs. 12.91 and 12.92 are not the most general transformation laws, for we began with a system $\mathcal{S}_0$ in which the charges were at rest and where, consequently, there was no magnetic field. To derive the *general* rule, we must start out in a system with both electric and magnetic fields. For this purpose $\mathcal{S}$ itself will serve nicely. In addition to the electric field

$$E_y = \frac{\sigma}{\epsilon_0}, \tag{12.94}$$

there is a *magnetic* field due to the surface currents (Fig. 12.35b):

$$\mathbf{K}_\pm = \mp\sigma v_0\,\hat{\mathbf{x}}. \tag{12.95}$$

By the right-hand rule, this field points in the negative z direction; its magnitude is given by Ampère's law (Ex. 5.8):

$$B_z = -\mu_0\sigma v_0. \tag{12.96}$$

In a *third* system, $\bar{\mathcal{S}}$, traveling to the right with speed v relative to $\mathcal{S}$ (Fig. 12.38), the fields would be

$$\bar{E}_y = \frac{\bar{\sigma}}{\epsilon_0}, \quad \bar{B}_z = -\mu_0\bar{\sigma}\bar{v}, \tag{12.97}$$

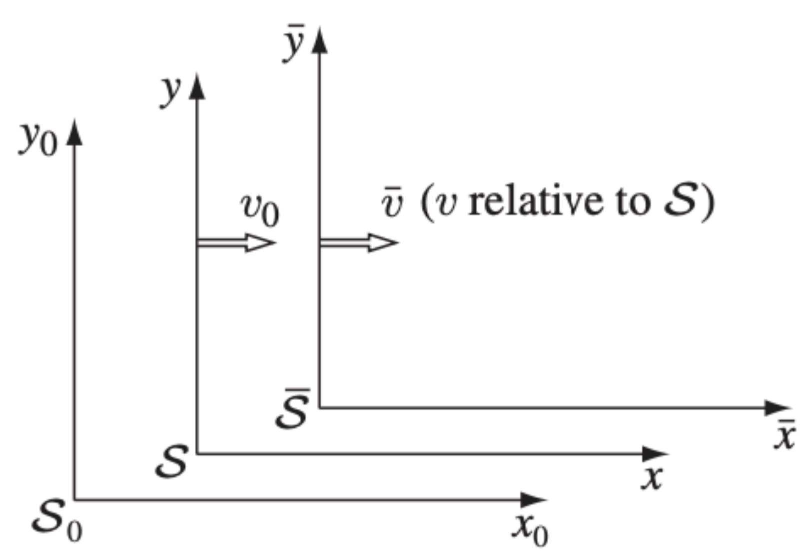

FIGURE 12.38

where $\bar{v}$ is the velocity of $\bar{\mathcal{S}}$ relative to $\mathcal{S}_0$:

$$\bar{v} = \frac{v + v_0}{1 + v v_0/c^2}, \quad \bar{\gamma} = \frac{1}{\sqrt{1 - \bar{v}^2/c^2}}, \tag{12.98}$$

and

$$\bar{\sigma} = \bar{\gamma}\sigma_0. \tag{12.99}$$

It remains only to express $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ (Eq. 12.97), in terms of $\mathbf{E}$ and $\mathbf{B}$ (Eqs. 12.94 and 12.96). In view of Eqs. 12.90 and 12.99, we have

$$\bar{E}_y = \left(\frac{\bar{\gamma}}{\gamma_0}\right)\frac{\sigma}{\epsilon_0}, \quad \bar{B}_z = -\left(\frac{\bar{\gamma}}{\gamma_0}\right)\mu_0\sigma\bar{v}. \tag{12.100}$$

With a little algebra, you can show that

$$\frac{\bar{\gamma}}{\gamma_0} = \frac{\sqrt{1 - v_0^2/c^2}}{\sqrt{1 - \bar{v}^2/c^2}} = \frac{1 + v v_0/c^2}{\sqrt{1 - v^2/c^2}} = \gamma\left(1 + \frac{v v_0}{c^2}\right), \tag{12.101}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \tag{12.102}$$

as always. Thus, writing $\bar{E}_y$ in terms of the components of $\mathbf{E}$ and $\mathbf{B}$ in $\mathcal{S}$,

$$\bar{E}_y = \gamma\left(1 + \frac{v v_0}{c^2}\right)\frac{\sigma}{\epsilon_0} = \gamma\left(E_y - \frac{v}{c^2\epsilon_0\mu_0}B_z\right),$$

whereas

$$\bar{B}_z = -\gamma\left(1 + \frac{v v_0}{c^2}\right)\mu_0\sigma\left(\frac{v + v_0}{1 + v v_0/c^2}\right) = \gamma(B_z - \mu_0\epsilon_0 v E_y).$$

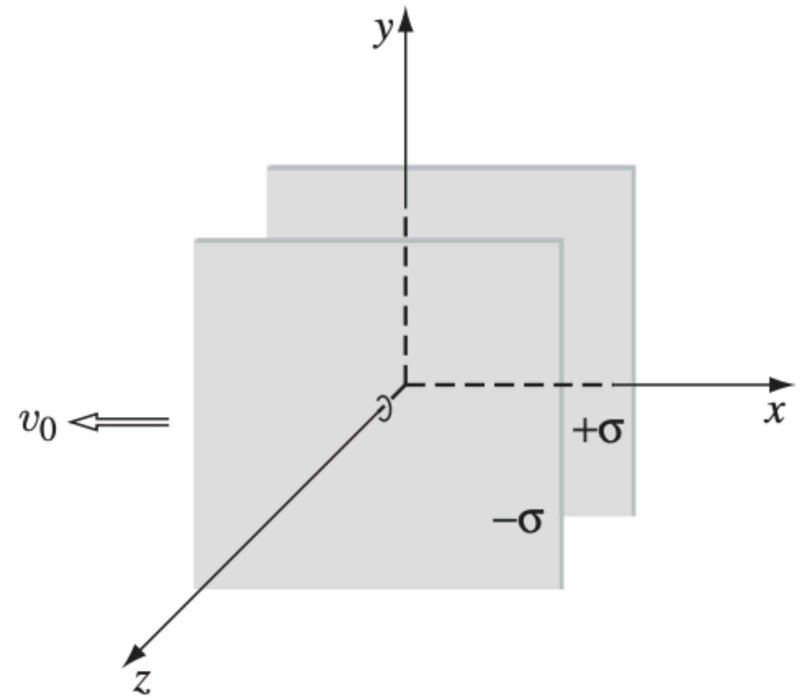

FIGURE 12.39

Or, since $\mu_0\epsilon_0 = 1/c^2$,

$$\left.\begin{aligned} \bar{E}_y &= \gamma(E_y - vB_z), \\ \bar{B}_z &= \gamma\left(B_z - \frac{v}{c^2}E_y\right). \end{aligned}\right\} \tag{12.103}$$

This tells us how E_y and B_z transform—to do E_z and B_y, we simply align the same capacitor parallel to the xy plane instead of the xz plane (Fig. 12.39). The fields in $\mathcal{S}$ are then

$$E_z = \frac{\sigma}{\epsilon_0}, \quad B_y = \mu_0 \sigma v_0.$$

(Use the right-hand rule to get the sign of B_y.) The rest of the argument is identical—everywhere we had E_y before, read E_z, and everywhere we had B_z, read $-B_y$:

$$\left.\begin{aligned} \bar{E}_z &= \gamma(E_z + vB_y), \\ \bar{B}_y &= \gamma\left(B_y + \frac{v}{c^2}E_z\right). \end{aligned}\right\} \tag{12.104}$$

As for the x components, we have already seen (by orienting the capacitor parallel to the yz plane) that

$$\bar{E}_x = E_x. \tag{12.105}$$

Since in this case there is no accompanying magnetic field, we cannot deduce the transformation rule for B_x. But another configuration will do the job: Imagine a long *solenoid* aligned parallel to the x axis (Fig. 12.40) and at rest in $\mathcal{S}$. The magnetic field within the coil is

$$B_x = \mu_0 n I, \tag{12.106}$$

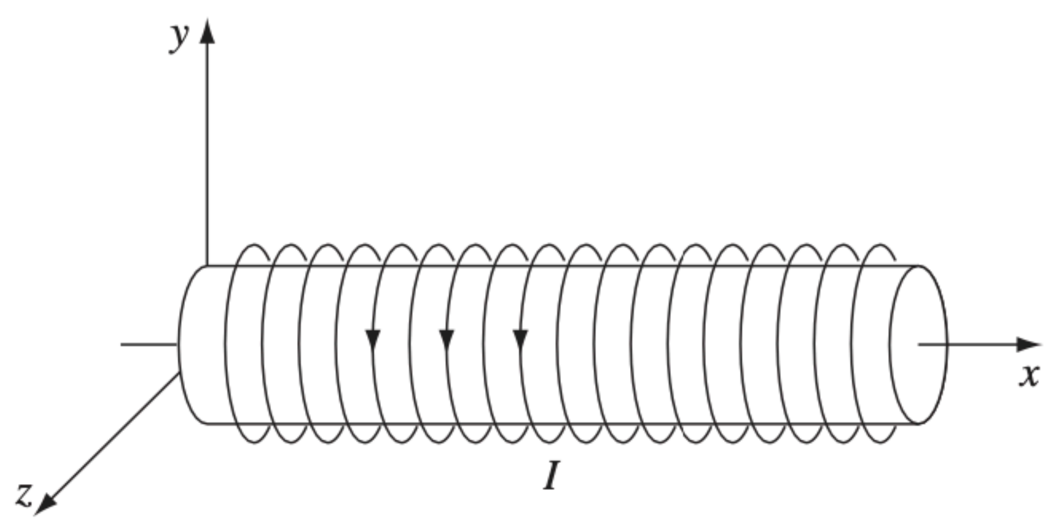

FIGURE 12.40

where n is the number of turns per unit length, and I is the current. In system $\bar{\mathcal{S}}$, the length contracts, so n *increases:*

$$\bar{n} = \gamma n. \tag{12.107}$$

On the other hand, time *dilates:* The $\mathcal{S}$ clock, which rides along with the solenoid, runs slow, so the current (charge per unit time) in $\bar{\mathcal{S}}$ is given by

$$\bar{I} = \frac{1}{\gamma} I. \tag{12.108}$$

The two factors of γ exactly cancel, and we conclude that

$$\bar{B}_x = B_x.$$

Like **E**, the component of **B** *parallel* to the motion is unchanged.

Here, then, is the complete set of transformation rules:

$$\boxed{\begin{array}{lll} \bar{E}_x = E_x, & \bar{E}_y = \gamma(E_y - vB_z), & \bar{E}_z = \gamma(E_z + vB_y), \\ \bar{B}_x = B_x, & \bar{B}_y = \gamma\left(B_y + \frac{v}{c^2}E_z\right), & \bar{B}_z = \gamma\left(B_z - \frac{v}{c^2}E_y\right). \end{array}} \tag{12.109}$$

Two special cases warrant particular attention:

1. If $\mathbf{B} = \mathbf{0}$ in $\mathcal{S}$, then

$$\bar{\mathbf{B}} = \gamma\frac{v}{c^2}(E_z\,\hat{\mathbf{y}} - E_y\,\hat{\mathbf{z}}) = \frac{v}{c^2}(\bar{E}_z\,\hat{\mathbf{y}} - \bar{E}_y\,\hat{\mathbf{z}}),$$

or, since $\mathbf{v} = v\,\hat{\mathbf{x}}$,

$$\boxed{\bar{\mathbf{B}} = -\frac{1}{c^2}(\mathbf{v} \times \bar{\mathbf{E}}).} \tag{12.110}$$

2. If $\mathbf{E} = \mathbf{0}$ in $\mathcal{S}$, then

$$\bar{\mathbf{E}} = -\gamma v(B_z\,\hat{\mathbf{y}} - B_y\,\hat{\mathbf{z}}) = -v(\bar{B}_z\,\hat{\mathbf{y}} - \bar{B}_y\,\hat{\mathbf{z}}),$$

or

$$\boxed{\bar{\mathbf{E}} = \mathbf{v} \times \bar{\mathbf{B}}.} \tag{12.111}$$

In other words, if either **E** or **B** is zero (at a particular point) in one system, then in any other system the fields (at that point) are very simply related by Eq. 12.110 or Eq. 12.111.

Example 12.15. Magnetic field of a point charge in uniform motion. Find the magnetic field of a point charge q moving at constant velocity $\mathbf{v}$.

Solution

In the particle's *rest* frame the magnetic field is zero (everywhere), so in a system moving with velocity $-\mathbf{v}$ (in which the *particle* is moving at velocity $+\mathbf{v}$)[24]

$$\mathbf{B} = \frac{1}{c^2}(\mathbf{v} \times \mathbf{E}).$$

We calculated the electric field in Ex. 12.14. The magnetic field, then, is

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{qv(1 - v^2/c^2)\sin\theta}{[1 - (v^2/c^2)\sin^2\theta]^{3/2}} \frac{\hat{\boldsymbol{\phi}}}{R^2}, \tag{12.112}$$

where $\hat{\boldsymbol{\phi}}$ aims counterclockwise as you face the oncoming charge. Incidentally, in the nonrelativistic limit ($v^2 \ll c^2$), Eq. 12.112 reduces to

$$\mathbf{B} \approx \frac{\mu_0}{4\pi} q \frac{\mathbf{v} \times \hat{\mathbf{R}}}{R^2},$$

which is exactly what you would get by naïve application of the Biot-Savart law to a point charge (Eq. 5.43).

12.3.3 ■ The Field Tensor

As Eq. 12.109 indicates, **E** and **B** certainly do *not* transform like the spatial parts of the two 4-vectors—in fact, the components of **E** and **B** are stirred together when you go from one inertial system to another. What sort of an object is this, which has six components and transforms according to Eq. 12.109? *Answer*: It's an **antisymmetric, second-rank tensor**.

Remember that a 4-vector transforms by the rule

$$\bar{a}^{\mu} = \Lambda^{\mu}_{\nu} a^{\nu} \tag{12.113}$$

(summation over ν implied), where Λ is the Lorentz transformation matrix. If $\bar{\mathcal{S}}$ is moving in the x direction at speed v, Λ has the form

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{12.114}$$

and Λ^{μ}_{ν} is the entry in row μ, column ν. A (second-rank) tensor is an object with *two* indices, which transforms with *two* factors of Λ (one for each index):

$$\bar{t}^{\mu\nu} = \Lambda^{\mu}_{\lambda} \Lambda^{\nu}_{\sigma} t^{\lambda\sigma}. \tag{12.115}$$

A tensor (in 4 dimensions) has $4 \times 4 = 16$ components, which we can display in a 4×4 array:

$$t^{\mu\nu} = \left\{ \begin{array}{cccc} t^{00} & t^{01} & t^{02} & t^{03} \\ t^{10} & t^{11} & t^{12} & t^{13} \\ t^{20} & t^{21} & t^{22} & t^{23} \\ t^{30} & t^{31} & t^{32} & t^{33} \end{array} \right\}.$$

However, the 16 elements need not all be different. For instance, a *symmetric* tensor has the property

$$t^{\mu\nu} = t^{\nu\mu} \quad (\textbf{symmetric tensor}). \tag{12.116}$$

In this case there are 10 distinct components; 6 of the 16 are repeats ($t^{01} = t^{10}$, $t^{02} = t^{20}$, $t^{03} = t^{30}$, $t^{12} = t^{21}$, $t^{13} = t^{31}$, $t^{23} = t^{32}$). Similarly, an *anti*symmetric tensor obeys

$$t^{\mu\nu} = -t^{\nu\mu} \quad (\textbf{antisymmetric tensor}). \tag{12.117}$$

Such an object has just 6 distinct elements—of the original 16, six are repeats (the same ones as before, only this time with a minus sign) and four are zero (t^{00}, t^{11}, t^{22}, and t^{33}). Thus, the general antisymmetric tensor has the form

$$t^{\mu\nu} = \left\{ \begin{array}{cccc} 0 & t^{01} & t^{02} & t^{03} \\ -t^{01} & 0 & t^{12} & t^{13} \\ -t^{02} & -t^{12} & 0 & t^{23} \\ -t^{03} & -t^{13} & -t^{23} & 0 \end{array} \right\}.$$

Let's see how the transformation rule (Eq. 12.115) works, for the six distinct components of an antisymmetric tensor. Starting with $\bar{t}^{01}$, we have

$$\bar{t}^{01} = \Lambda^0_\lambda \Lambda^1_\sigma t^{\lambda\sigma},$$

but according to Eq. 12.114, $\Lambda^0_\lambda = 0$ unless $\lambda = 0$ or 1, and $\Lambda^1_\sigma = 0$ unless $\sigma = 0$ or 1. So there are four terms in the sum:

$$\bar{t}^{01} = \Lambda^0_0 \Lambda^1_0 t^{00} + \Lambda^0_0 \Lambda^1_1 t^{01} + \Lambda^0_1 \Lambda^1_0 t^{10} + \Lambda^0_1 \Lambda^1_1 t^{11}.$$

On the other hand, $t^{00} = t^{11} = 0$, while $t^{01} = -t^{10}$, so

$$\bar{t}^{01} = (\Lambda^0_0\Lambda^1_1 - \Lambda^0_1\Lambda^1_0)t^{01} = (\gamma^2 - (\gamma\beta)^2)t^{01} = t^{01}.$$

I'll let you work out the others—the complete set of transformation rules is

$$\left.\begin{array}{lll} \bar{t}^{01} = t^{01}, & \bar{t}^{02} = \gamma(t^{02} - \beta t^{12}), & \bar{t}^{03} = \gamma(t^{03} + \beta t^{31}), \\ \bar{t}^{23} = t^{23}, & \bar{t}^{31} = \gamma(t^{31} + \beta t^{03}), & \bar{t}^{12} = \gamma(t^{12} - \beta t^{02}). \end{array}\right\} \tag{12.118}$$

These are precisely the rules we obtained on physical grounds for the electromagnetic fields (Eq. 12.109)—in fact, we can construct the **field tensor** $F^{\mu\nu}$ by direct comparison:[25]

$$F^{01} \equiv \frac{E_x}{c}, \quad F^{02} \equiv \frac{E_y}{c}, \quad F^{03} \equiv \frac{E_z}{c}, \quad F^{12} \equiv B_z, \quad F^{31} \equiv B_y, \quad F^{23} \equiv B_x.$$

Written as an array,

$$F^{\mu\nu} = \left\{\begin{array}{cccc} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{array}\right\}. \tag{12.119}$$

Thus relativity completes and perfects the job begun by Oersted, combining the electric and magnetic fields into a single entity, $F^{\mu\nu}$.

If you followed that argument with exquisite care, you may have noticed that there was a different way of imbedding **E** and **B** in an antisymmetric tensor: Instead of comparing the first line of Eq. 12.109 with the first line of Eq. 12.118, and the second with the second, we could relate the first line of Eq. 12.109 to the *second* line of Eq. 12.118, and vice versa. This leads to **dual tensor,** $G^{\mu\nu}$:

$$G^{\mu\nu} = \left\{ \begin{array}{cccc} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z/c & E_y/c \\ -B_y & E_z/c & 0 & -E_x/c \\ -B_z & -E_y/c & E_x/c & 0 \end{array} \right\}. \tag{12.120}$$

$G^{\mu\nu}$ can be obtained directly from $F^{\mu\nu}$ by the substitution $\mathbf{E}/c \to \mathbf{B}$, $\mathbf{B} \to -\mathbf{E}/c$. Notice that this operation leaves Eq. 12.109 unchanged—that's why both tensors generate the correct transformation rules for **E** and **B.**

12.3.4 ■ Electrodynamics in Tensor Notation

Now that we know how to represent the fields in relativistic notation, it is time to reformulate the laws of electrodynamics (Maxwell's equations and the Lorentz force law) in that language. To begin with, we must determine how the *sources* of the fields, ρ and **J**, transform. Imagine a cloud of charge drifting by; we concentrate on an infinitesimal volume V, which contains charge Q moving at velocity **u** (Fig. 12.43). The charge density is

$$\rho = \frac{Q}{V},$$

and the current density[26] is

$$\mathbf{J} = \rho\mathbf{u}.$$

I would like to express these quantities in terms of the **proper charge density** ρ_0, the density *in the rest system of the charge:*

$$\rho_0 = \frac{Q}{V_0},$$

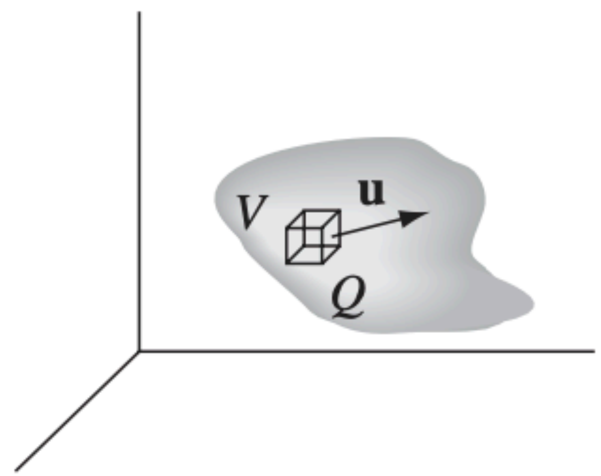

FIGURE 12.43

where V_0 is the rest volume of the cloud. Because one dimension (the one along the direction of motion) is Lorentz-contracted,

$$V = \sqrt{1 - u^2/c^2}\, V_0, \tag{12.121}$$

and hence

$$\rho = \rho_0 \frac{1}{\sqrt{1 - u^2/c^2}}, \quad \mathbf{J} = \rho_0 \frac{\mathbf{u}}{\sqrt{1 - u^2/c^2}}. \tag{12.122}$$

Comparing this with Eqs. 12.40 and 12.42, we recognize here the components of proper velocity, multiplied by the invariant ρ_0. Evidently charge density and current density go together to make a 4-vector:

$$J^\mu = \rho_0 \eta^\mu, \tag{12.123}$$

whose components are

$$\boxed{J^\mu = (c\rho, J_x, J_y, J_z).} \tag{12.124}$$

We'll call it the **current density 4-vector.**

The continuity equation (Eq. 5.29),

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t},$$

expressing the local conservation of charge, takes on a nice compact form when written in terms of J^μ. For

$$\nabla \cdot \mathbf{J} = \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} = \sum_{i=1}^{3} \frac{\partial J^i}{\partial x^i},$$

while

$$\frac{\partial \rho}{\partial t} = \frac{1}{c}\frac{\partial J^0}{\partial t} = \frac{\partial J^0}{\partial x^0}. \tag{12.125}$$

Thus, bringing $\partial\rho/\partial t$ over to the left side (in the continuity equation), we have:

$$\boxed{\frac{\partial J^\mu}{\partial x^\mu} = 0,} \tag{12.126}$$

with summation over μ implied. Incidentally, $\partial J^\mu/\partial x^\mu$ is the four-dimensional *divergence* of J^μ, so the continuity equation states that the current density 4-vector is divergenceless.

As for Maxwell's equations, they can be written

$$\boxed{\frac{\partial F^{\mu\nu}}{\partial x^\nu} = \mu_0 J^\mu, \quad \frac{\partial G^{\mu\nu}}{\partial x^\nu} = 0,} \tag{12.127}$$

with summation over ν implied. Each of these stands for four equations—one for every value of μ. If $\mu = 0$, the first equation reads

$$\frac{\partial F^{0\nu}}{\partial x^\nu} = \frac{\partial F^{00}}{\partial x^0} + \frac{\partial F^{01}}{\partial x^1} + \frac{\partial F^{02}}{\partial x^2} + \frac{\partial F^{03}}{\partial x^3}$$

$$= \frac{1}{c}\left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right) = \frac{1}{c}(\boldsymbol{\nabla} \cdot \mathbf{E})$$

$$= \mu_0 J^0 = \mu_0 c\rho,$$

or

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho.$$

This, of course, is Gauss's law. If $\mu = 1$, we have

$$\frac{\partial F^{1\nu}}{\partial x^\nu} = \frac{\partial F^{10}}{\partial x^0} + \frac{\partial F^{11}}{\partial x^1} + \frac{\partial F^{12}}{\partial x^2} + \frac{\partial F^{13}}{\partial x^3}$$

$$= -\frac{1}{c^2}\frac{\partial E_x}{\partial t} + \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = \left(-\frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B}\right)_x$$

$$= \mu_0 J^1 = \mu_0 J_x.$$

Combining this with the corresponding results for $\mu = 2$ and $\mu = 3$ gives

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t},$$

which is Ampère's law with Maxwell's correction.

Meanwhile, the second equation in 12.127, with $\mu = 0$, becomes

$$\frac{\partial G^{0\nu}}{\partial x^\nu} = \frac{\partial G^{00}}{\partial x^0} + \frac{\partial G^{01}}{\partial x^1} + \frac{\partial G^{02}}{\partial x^2} + \frac{\partial G^{03}}{\partial x^3}$$

$$= \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = \nabla \cdot \mathbf{B} = 0$$

(the third of Maxwell's equations), whereas $\mu = 1$ yields

$$\frac{\partial G^{1\nu}}{\partial x^\nu} = \frac{\partial G^{10}}{\partial x^0} + \frac{\partial G^{11}}{\partial x^1} + \frac{\partial G^{12}}{\partial x^2} + \frac{\partial G^{13}}{\partial x^3}$$

$$= -\frac{1}{c}\frac{\partial B_x}{\partial t} - \frac{1}{c}\frac{\partial E_z}{\partial y} + \frac{1}{c}\frac{\partial E_y}{\partial z} = -\frac{1}{c}\left(\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E}\right)_x = 0.$$

So, combining this with the corresponding results for $\mu = 2$ and $\mu = 3$,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t},$$

which is Faraday's law. In relativistic notation, then, Maxwell's four rather cumbersome equations reduce to two delightfully simple ones.

In terms of $F^{\mu\nu}$ and the proper velocity η^μ, the Minkowski force on a charge q is given by

$$\boxed{K^\mu = q\eta_\nu F^{\mu\nu}.} \tag{12.128}$$

For if $\mu = 1$, we have

$$K^1 = q\eta_\nu F^{1\nu} = q(-\eta^0 F^{10} + \eta^1 F^{11} + \eta^2 F^{12} + \eta^3 F^{13})$$

$$= q\left[\frac{-c}{\sqrt{1 - u^2/c^2}}\left(\frac{-E_x}{c}\right) + \frac{u_y}{\sqrt{1 - u^2/c^2}}(B_z) + \frac{u_z}{\sqrt{1 - u^2/c^2}}(-B_y)\right]$$

$$= \frac{q}{\sqrt{1 - u^2/c^2}}\,[\mathbf{E} + (\mathbf{u} \times \mathbf{B})]_x,$$

with a similar formula for $\mu = 2$ and $\mu = 3$. Thus,

$$\mathbf{K} = \frac{q}{\sqrt{1 - u^2/c^2}} [\mathbf{E} + (\mathbf{u} \times \mathbf{B})], \tag{12.129}$$

and therefore, referring back to Eq. 12.69,

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

which is the Lorentz force law. Equation 12.128, then, represents the Lorentz force law in relativistic notation. I'll leave for you the interpretation of the zeroth component (Prob. 12.55).

12.3.5 ■ Relativistic Potentials

From Chapter 10, we know that the electric and magnetic fields can be expressed in terms of a scalar potential V and a vector potential $\mathbf{A}$:

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{B} = \nabla \times \mathbf{A}. \tag{12.131}$$

As you might guess, V and $\mathbf{A}$ together constitute a 4-vector:

$$\boxed{A^\mu = (V/c, A_x, A_y, A_z).} \tag{12.132}$$

In terms of this **4-vector potential**, the field tensor can be written

$$\boxed{F^{\mu\nu} = \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu}.} \tag{12.133}$$

(Observe that the differentiation is with respect to the *covariant* vectors x_μ and x_ν; remember, that changes the sign of the zeroth component: $x_0 = -x^0$. See Prob. 12.56.)

To check that Eq. 12.133 is equivalent to Eq. 12.131, let's evaluate a few terms explicitly. For $\mu = 0$, $\nu = 1$,

$$F^{01} = \frac{\partial A^1}{\partial x_0} - \frac{\partial A^0}{\partial x_1} = -\frac{\partial A_x}{\partial (ct)} - \frac{1}{c}\frac{\partial V}{\partial x}$$

$$= -\frac{1}{c}\left(\frac{\partial \mathbf{A}}{\partial t} + \nabla V\right)_x = \frac{E_x}{c}.$$

That (and its companions with $\nu = 2$ and $\nu = 3$) is the first equation in Eq. 12.131. For $\mu = 1$, $\nu = 2$, we get

$$F^{12} = \frac{\partial A^2}{\partial x_1} - \frac{\partial A^1}{\partial x_2} = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = (\nabla \times \mathbf{A})_z = B_z,$$

which (together with the corresponding results for F^{23} and F^{31}) is the second equation in Eq. 12.131.

The potential formulation automatically takes care of the homogeneous Maxwell equation ($\partial G^{\mu\nu}/\partial x^\nu = 0$). As for the inhomogeneous equation ($\partial F^{\mu\nu}/\partial x^\nu = \mu_0 J^\mu$), that becomes

$$\frac{\partial}{\partial x_\mu}\left(\frac{\partial A^\nu}{\partial x^\nu}\right) - \frac{\partial}{\partial x_\nu}\left(\frac{\partial A^\mu}{\partial x^\nu}\right) = \mu_0 J^\mu. \qquad (12.134)$$

This is an intractable equation as it stands. However, you will recall that the potentials are not uniquely determined by the fields—in fact, it's clear from Eq. 12.133 that you could add to A^μ the gradient of any scalar function λ:

$$A^\mu \longrightarrow A^{\mu\prime} = A^\mu + \frac{\partial \lambda}{\partial x_\mu}, \tag{12.135}$$

without changing $F^{\mu\nu}$. This is precisely the **gauge invariance** we noted in Chapter 10; we can exploit it to simplify Eq. 12.134. In particular, the Lorenz gauge condition (Eq. 10.12)

$$\boldsymbol{\nabla} \cdot \mathbf{A} = -\frac{1}{c^2}\frac{\partial V}{\partial t}$$

becomes, in relativistic notation,

$$\frac{\partial A^\mu}{\partial x^\mu} = 0. \tag{12.136}$$

In the Lorenz gauge, therefore, Eq. 12.134 reduces to

$$\boxed{\Box^2 A^\mu = -\mu_0 J^\mu,} \tag{12.137}$$

where $\Box^2$ is the **d'Alembertian**,

$$\Box^2 \equiv \frac{\partial}{\partial x_\nu}\frac{\partial}{\partial x^\nu} = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}. \tag{12.138}$$

Equation 12.137 combines our previous results into a single 4-vector equation—it represents the most elegant formulation of Maxwell's equations.[27]