June 1996

Revised March 2010

Introduction to Special Relativity

Paul Renteln

Department of Physics
California State University
5500 University Parkway
San Bernardino, CA 92407

Table of Contents

1. Introduction

1.1 In 1905 Einstein revolutionized our understanding of the physical universe with a series of three remarkable papers in *Annalen der Physik*. The first two successfully explained two outstanding problems of the day, Brownian motion and the photoelectric effect. The latter paper established without doubt the validity of the particulate nature of light, and helped usher in the age of quantum theory.

The last of the three papers of 1905, entitled “On the Electrodynamics of Moving Bodies”, was so radical in nature and so extensive in scope that it met, at first, with some resistance from the physics community. But it soon became apparent that the theory put forth in this paper, which we now call the theory of special relativity (or the special theory of relativity), was completely consistent and entirely correct. The extraordinary implications of this theory, including the idea that time and space themselves do not have absolute meaning, shattered the complacency of those steeped in 18th century absolutism and helped spawn intellectual revolutions in the most tangential disciplines. For better or worse, Einstein's theory of special relativity allowed mankind to tap the awesome energies of the nucleus. More than anything else, however, his theory revealed some of the fascinating subtleties of nature, and led to a more profound understanding of the universe itself.

2. Reference Frames

2.1 Before we can begin our study of special relativity we need to agree on the meaning of certain key concepts. So we make the following

Definition. An **event** is a happening at some point of space at some instant of time.

For example, the explosion of a firecracker at the origin at time $t = 0$ is an event, but a sale at Bullock's is not, because it does not take place at a specific instant of time.

Definition. **Spacetime** is the collection of all possible events.

Definition. A **reference frame** is a spacetime grid that locates each event by specifying its (t, x, y, z) coordinates.

2.2 Example. In many of our examples we shall employ two reference frames, one stuck to the Earth, called the **lab frame**, and one attached to a rocket, called the **rocket frame** (see Figure 1).

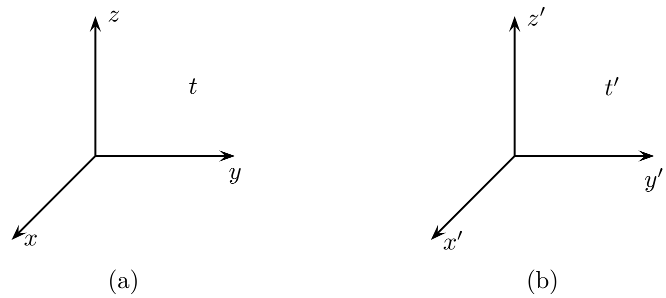


Figure 1. (a) Earth based (laboratory) reference frame with grid stuck to the Earth. (b) Rocket reference frame with grid stuck to rocket.

2.3 For the most part, we shall restrict ourselves to motion in one dimension, in which case our reference frames will be in $1+1$ dimensions (1 spatial and 1 temporal dimension) rather than $3+1$ dimensions, as above.

Next, recall that **Newton's Law of Inertia** states that if the net force on a body is zero, it moves with constant velocity. This allows us to make the following

Definition. An **inertial reference frame** is a reference frame in which Newton's Law of Inertia is valid. An observer for whom the inertial reference frame is at rest is called an **inertial observer**.

2.4 Example. Suppose an ice skater is gliding across the ice (see Figure 2). The same ice skater is viewed by four different observers, who record her motion relative to their reference frames. The first three are inertial observers, while the last is a non-inertial observer, because this observer sees the skater move along a curved path despite the fact that no forces act on her.

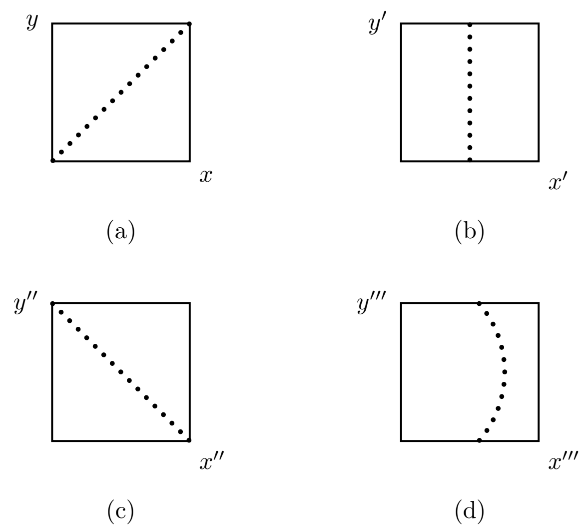


Figure 2. (a) Trajectory of ice skater as seen by person in bleachers (at rest in lab frame). (b) Trajectory of ice skater as seen by person running to right with constant horizontal speed equal to that of the skater. (c) Trajector of ice skater as seen by person running to right with constant horizontal speed greater than that of the skater. (d) Trajectory of ice skater as seen by person accelerating to the right. This observer is non-inertial.

3. The Postulates of Special Relativity

3.1 Einstein derived his theory of special relativity by thinking deeply about some of the implications of James Clerk Maxwell's theory of electrodynamics (hence the title of Einstein's paper of 1905). Consider a bar magnet and a loop of wire, as shown in Figure 3. We know that if we move the bar magnet toward the loop of wire, there will be a current induced in the wire. This was demonstrated convincingly by Michael Faraday around 1840. Similarly, if we move the loop of wire toward the bar magnet (at the same relative rate as before) an identical current will be induced in the wire. All this was well known 50 years before Einstein.

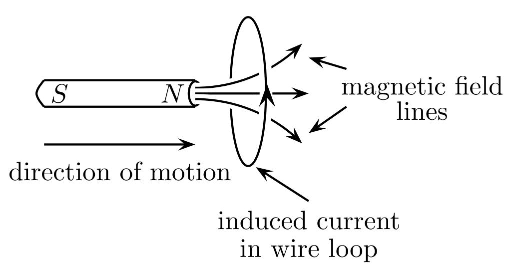


Figure 3. A bar magnet is moved toward a loop of wire, thereby inducing a current in the wire.

3.2 What bothered Einstein was that Maxwell's theory provided **two distinct explanations** for what really ought to be thought of as a single phenomenon. In the case in which the wire is moving and the magnet is stationary, the explanation for the current is that the electrons in the wire are moving in a fixed magnetic field, and hence experience a force (given by $\boldsymbol{F} = q\boldsymbol{v} \times \boldsymbol{B}$). This **magnetic** force causes them to move, whence a current arises.

On the other hand, when the magnet is moving and the wire is stationary, the electrons in the wire do not experience a magnetic force (because they are not moving). Rather what happens is that the electrons feel a time varying magnetic field, which produces an electric field in their vicinity, and hence they experience an **electric** force (given by $\boldsymbol{F} = q\boldsymbol{E}$). It is this electric force that gives rise to the current in the loop.

3.3 Einstein felt this situation was unaesthetic: one phenomenon but two theoretical explanations. It seemed to him that one ought to be able to cast Maxwell's theory in a form that showed no preference for what really only amounted to a different point of view. In other words, the laws of physics ought to be the same for everyone, regardless of his reference frame.

Of course, Einstein was well aware that, in some cases, this condition could not be fulfilled. For example, it is clear from the example of the ice skater that, for the non-inertial observer, Newton's Law of Inertia does not seem to hold. Although Einstein was to eventually construct his theory of general relativity to incorporate non-inertial observers, at this point in the game he was only interested in inertial observers, and relative motion at constant velocity. So he put forward the following quite reasonable postulate: [1]

Postulate 1. The laws of physics are the same in every inertial reference frame.

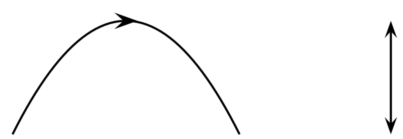
3.4 If you think about it, this postulate is certainly reasonable. For example, it is true that, if you are in an airplane moving at 600 mph, your teacup does not suddenly jump out of your hands (unless you hit some turbulence; but then you are being accelerated and are no longer in an inertial reference frame). That is, Newton's Law of Inertia still holds. And if you apply a force to an object and measure its acceleration, you still deduce that

[1] This idea is actually a very old one. It was Galileo who first put forward an idea like this, some 300 years before Einstein. But it took Einstein's genius to use this as one of the foundations of a new physical theory.

$F = ma$. Maxwell's equations of electrodynamics hold equally well on the airplane (all the electrical systems still work, do they not?) as on the ground.

3.5 It should be remarked that postulate 1 does not assert that every inertial observer will observe the same phenomenon in exactly the same way; this is certainly false (*cf.* Figure 2). What it says is that while the different observers may observe the same phenomenon in different ways, they will all deduce the same underlying **laws**.

3.6 Example. Suppose I hit a baseball. The people in the bleachers see something a little different from the player running beneath the ball to catch it, but they both agree on the physical law, namely that distance equals velocity times time.



(a) (b)

Figure 4. (a) The trajectory of the baseball as seen by a person in the bleachers. This person sees the baseball move forward with $v_x \neq 0$. According to him, it travels a distance $x = v_x t$. (b) The trajectory of the baseball as seen by a person running underneath to catch it. This person sees no forward motion for the baseball ($v'_x = 0$). According to him, the ball moves a horizontal distance given by $x' = v'_x t$. Both observers deduce the *same* kinematical law.

3.7 Postulate 1 makes a very strong statement. It implies that there is no such thing as an **absolute velocity**. That is, it becomes impossible, **in principle**, to determine whether one is moving at constant velocity or whether one is at rest in some reference frame. So the notion of absolute velocity is meaningless, because it cannot be measured. This goes against hundreds of years of physical intuition in which people, following Newton, assumed that some sense could be made of the idea of **absolute rest**. We now know, for example, that our entire galaxy is moving in the direction of the constellation Virgo at some enormous rate of speed. But you certainly cannot not tell, can you?

3.8 The second idea presented itself to Einstein in the form of a question (a question that he first thought of as a teenager). He asked himself what it would be like to ride on a light beam. Would you see light, or would you see nothing? If you saw light, how fast would it be moving? Appealing to Maxwell's theory of electrodynamics once again, Einstein formulated the following answer. According to Maxwell's theory, light consists of electromagnetic waves traveling through the **aether**. [2] Now, by performing experiments on charged bodies in my reference frame I can deduce Maxwell's equations. From these equations I can then **deduce** the speed of light from the permittivity of free space ϵ_0 and the permeability of free space μ_0, according to the formula $c = (\epsilon_0\mu_0)^{-1/2}$. If I do this, I get $c = 3 \times 10^8$ m/s.

3.9 Einstein probably reasoned as follows. Suppose I shine a light beam in some direction. As measured in the laboratory frame, the beam moves at the speed of light. Now fly along side the light beam. According to Maxwell's theory, I ought to again measure the same speed for the light beam, namely 3×10^8 m/s. Of course, it does not have to be this way. That is, it is possible that there is a difference between light that **originates** in one frame and light that already exists but is **measured** from a moving frame. This would not necessarily conflict with the first postulate, because the **laws** of physics (*e.g.,* Maxwell's equations) would remain the same. But it certainly seems implausible. Because of this, Einstein introduced his second postulate:

Postulate 2. The speed of light is independent of the motion of the source.

3.10 This postulate resolves the question of how fast the light beam travels to an observer riding along with it—it still moves at speed c. [3] This is also consistent with Maxwell's equations, and their validity in all frames of reference.

3.11 The problem, however, is that Postulate 2 is in conflict with our common sense. For example, it seems intuitively obvious that, if you are walking at 3 mph inside a train that is moving forward at 5 mph, then relative to the ground you are moving at 8 mph. According to Postulate 2, however, if I measure the speed of light coming from a flashlight moving forward at, say, 2000 m/s, **I would still measure the speed to be** 3×10^8 **m/s**!

[2] Aether is a hypothetical substance first introduced by the ancient Greeks to explain the motion of the heavenly bodies. If aether permeated the universe, we could measure our velocity relative to it, and hence define absolute velocities. As we cannot do this (this is the result of the famous Michelson-Morley experiment—see Footnote 8) the aether is unobservable, hence unphysical.

[3] Actually, it turns out to be impossible to move that fast. But if you move close to the speed of light, the light beam still travels at the speed c.

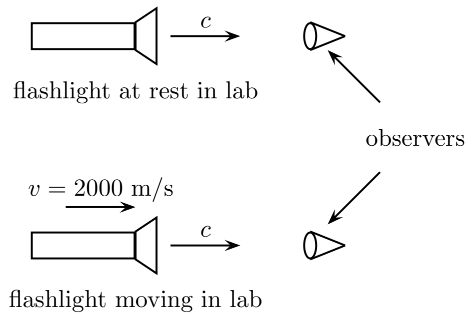


Figure 5. Both observers measure the tip of the light beam to moving at speed c relative to the laboratory.

How is this possible? Well, you might imagine that somehow things are just different for light beams. Or you might imagine that we cannot measure the speed of light so precisely, so that the answer might very well be $3 \times 10^8 + 2000$ m/s. But in both cases you would be mistaken. It is easy to dispose of the last objection: years of precise experimentation have convinced us that Postulate 2 is correct. The first objection is wrong, but to see why we must investigate a little more deeply.

4. Consequences of the Relativity Postulates

4.1 One immediate implication of Einstein's two postulates is that we must radically revise our intuitive notions of space and time. Our intuitions tell us that space and time themselves are immutable: a meter is always a meter, a second is always a second. Certainly this was the commonsense view that dominated 2000 years of physics. As we will see, the theory of special relativity forces us to give up these comfortable notions. We are forced to conclude that space and time themselves depend on one's state of motion!

Time Dilation

4.2 We begin by considering a **light clock** which is a device for measuring time using light flashes. The light clock works by sending a flash of light toward a mirror, and recording when it returns. As the speed of light is constant, the light flash ought to always return in the same amount of time. In our case, this time is simply $2D/c$. We use this time as our

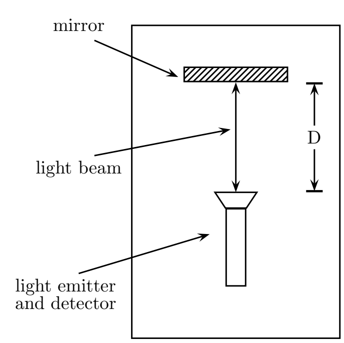


Figure 6. A light clock.

standard. [4] If, as we agreed, the speed of light were a universal constant, independent of the motion of the source, then this clock should work equally well in any inertial reference frame. We could calibrate all our clocks and watches to it, and use it to measure the rate of time's passage.

4.3 Now, imagine a train (for ease of illustration, a short train) in which we have put light clock. Let us call the train the **rocket frame** (for it is equipped with rocket engines and is moving very fast) and the Earth the **lab frame**. We then cause the light clock to flash once. ("One ping only, Vasily" [5]). In the rocket frame (that is, to someone in the same compartment as the light clock), the sequence of events is the same as always: the light emerges from the source, then travels to the mirror and back.

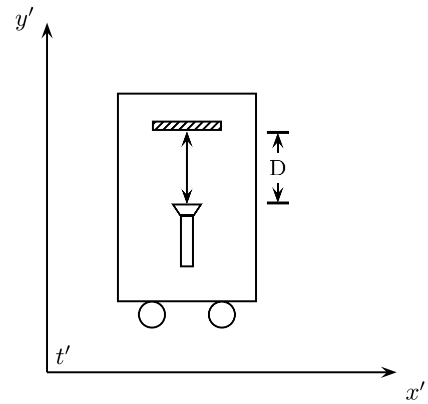


Figure 7. Trajectory of the light beam as seen by the rocket observer.

4.4 According to the rocket observer, the two events (1) emission of the light signal, and (2) receipt of the light signal, take place at the same point in space but at times separated by $2D/c$. To make things precise, let us agree to use primed spacetime coordinates for the rocket frame, and unprimed spacetime coordinates for the lab frame. So, for example, the time of emission in the rocket is denoted t'_1, while in the lab frame it is denoted t_1; similarly, the position of event 1 is x'_1 in the rocket frame and x_1 in the lab frame, and so forth. Then the rocket observer records the following:

$$\Delta x' = x'_2 - x'_1 = 0 \tag{4.1}$$

[4] Actually, the second is defined by means of the frequency of the radiation emitted by a certain transition of the cesium atom. The meter is defined as the distance traveled by light in 1/299,792,458 of a second.

[5] From "The Hunt for Red October", starring Sean Connery and Alec Baldwin.

and

$$\Delta t' = t_2' - t_1' = 2\frac{D}{c} \tag{4.2}$$

4.5 On the other hand, the same sequence of events is viewed rather differently from the perspective of the ground based observer. He sees the following trajectory for the light signal as it travels from emission point to reception point:

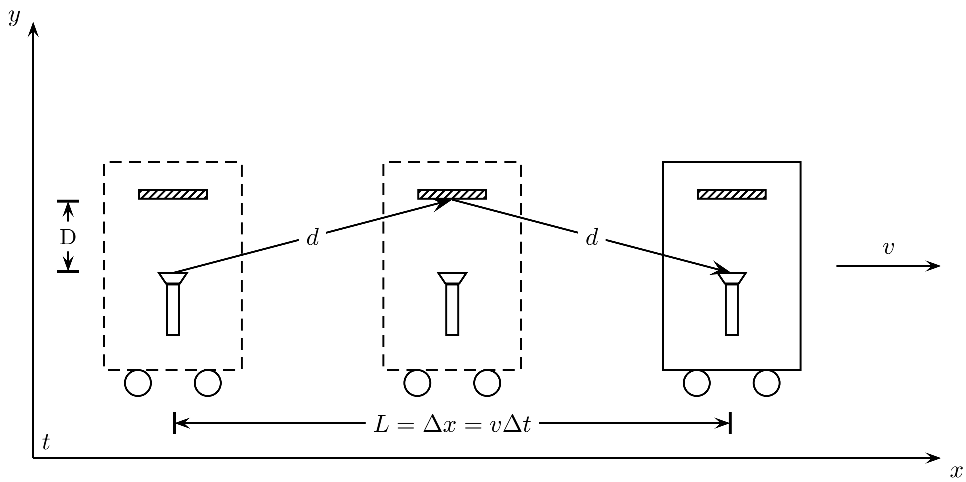


Figure 8. Trajectory of the light flash as seen by the laboratory observer.

According to the lab observer, the values of the temporal and spatial distances between the two events (emission and reception of the light) are different from those recorded by the rocket observer. Of course, this may not strike you as remarkable. “Of course things look different from the perspective of the lab observer. So what? We expect that.”

4.6 The problem is not that things **look** different, they **are** different. We may see this as follows. From elementary geometry we can immediately conclude that the distance and time between events 1 and 2, as measured by the lab observer, are given by

$$L := \Delta x = x_2 - x_1 = v\,\Delta t \tag{4.3}$$

and

$$\Delta t = t_2 - t_1 = 2\frac{d}{c} = \frac{2}{c}\sqrt{D^2 + (L/2)^2} \tag{4.4}$$

respectively.

4.7 Again you may ask, “so what”? Well, the point is that $\Delta t > \Delta t'$ (as one can immediately see by comparing equations (4.2) and (4.4)), so the lab observer, measuring the length of a 'tick' of the rocket observer's clock, observes it to be greater than expected. That is, **according to the lab observer, the rocket clock is running slow**.

4.8 The consequences of this inescapable conclusion are startling. Because it is not simply that the light clock is running slow, **time itself is running slow!** After all, all the clocks on the train are synchronized with the light clock: the conductor's watch, the pendulum clock, the biological clocks of the people on the train, and so on. If one of them is measured to be running slow, all of them are. According to the lab observer, everything on the train is occurring as if in slow motion!

4.9 How did this bizarre state of affairs come to pass? Reviewing the above derivation, it is clear that the essential point that led to this odd conclusion was that **the speed of light is the same for both observers**. Consider, for a moment, how the analogous argument would go with a monkey clock instead of a light clock. That is, suppose a monkey were climbing up and down a rope (at constant speed!) in the train, and that we used this setup as a clock. According to Newtonian physics, the lab observer would now measure exactly the same time as the rocket observer, because, relative to the ground, the monkey is moving **faster** than it is relative to the train. [6]

4.10 "Well", you say, "this explains it. Something strange must be happening with light." But this explanation will not work. Light clocks are one of the best kinds of clocks we have. They keep perfect time. If the monkey clock and the light clock were to become unsynchronized in the lab frame while remaining synchronized in the rocket frame, we could identify a preferred rest frame (*i.e.,* define an absolute velocity). But every experiment we have ever performed tells us this is impossible. This must mean that our intuitions about monkeys are mistaken, and that the Newtonian result is wrong. We are therefore forced to accept the conclusion that **time passes at different rates for observers with differing states of motion**.

4.11 Let us make the analysis a little more quantitative. Squaring (4.4) we get

$$(\Delta t)^2 = \frac{4D^2}{c^2} + \frac{4}{c^2}\frac{L^2}{4} = \frac{4D^2}{c^2} + \frac{(v\Delta t)^2}{c^2}$$

[6] For example, on the upward leg, the lab observer **computes, using Newtonian theory** the monkey's speed to be $v = \sqrt{v_x^2 + v_y^2} > v_y$. (In fact, this calculation is incorrect, as Newtonian theory is superseded by the special theory of relativity.)

where we used (4.3). Employing (4.2) then gives

$$(\Delta t)^2 = (\Delta t')^2 + \left(\frac{v}{c}\right)^2 (\Delta t)^2$$

which implies

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

or

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - (v/c)^2}} \tag{4.5}$$

4.12 Because the same symbols arise frequently in the study of special relativity, we introduce some shorthand notation. We define

$$\boxed{\beta := \frac{v}{c}} \tag{4.6}$$

representing the ratio of the speed of a body to the speed of light, and

$$\boxed{\gamma := \frac{1}{\sqrt{1 - \beta^2}}} \tag{4.7}$$

which is sometimes called the **gamma factor** or the **dilation factor**. It is worth pointing out that, as we shall later learn, the speed v of any object can never exceed the speed of light c, so β is always between 0 and 1, and γ is always greater than one. Also, observe that both numbers are dimensionless. Finally, as a matter of terminology, we sometimes refer to β as the speed of the object, when we really mean $v = \beta c$.

4.13 γ is called the dilation factor because we can write (4.5) as

$$\boxed{\Delta t = \gamma \Delta t'} \tag{4.8}$$

in which case it becomes apparent that γ indeed represents the degree to which time appears to be dilated (expanded or lengthened) according to the lab observer.

4.14 Example. Suppose the rocketship carrying the light clock is moving past the Earth observer with a speed $v = 2.12 \times 10^8$ m/s. Then we have $\beta = 0.71$ and $\gamma = 1.41$ so $\Delta t = 1.41\ \Delta t'$. This means that the Earth observer sees the light clock running **slow** by a factor of 1.41 (*i.e.,* one "clock-tick" on the rocket = 1 rocket second = $\Delta t' = 1$ s implies $\Delta t = 1.41$ s).

4.15 "If all this is true", you ask, "why do I never see you moving in slow motion when you run past me?" The answer, of course, is that the speeds of most objects around us are so small compared to the speed of light that we cannot notice these effects.

4.16 Example. Consider a typical jetliner that flies at a speed of 500 mph. This corresponds to a value for β equal to 7.5×10^{-7}. As β is so small, we may compute γ using a Taylor series expansion:

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} = (1-\beta^2)^{-1/2} \approx 1 + \beta^2/2 \tag{4.9}$$

which means, in this case, $\gamma \approx 1 + 2.8 \times 10^{-13}$. Hence

$$\Delta t = \Delta t'(1 + 2.8 \times 10^{-13})$$

and so

$$\frac{\Delta t - \Delta t'}{\Delta t'} = 2.8 \times 10^{-13}$$

In order for the two clocks to differ by 1 second, 110,000 years must pass on the rocket! Despite the smallness of this effect, **it was observed in 1971** by J.C. Hafele and R.E. Keating (*Science*, **177** (1972) 166). Messrs. Hafele and Keating flew four portable atomic clocks around the world on a jet plane, and when they were compared to the reference clock at the US Naval Observatory in Washington, D.C., it was discovered that the flying clocks differed from the naval clocks by around 275 nanoseconds, in perfect agreement with the special relativistic prediction. [7]

4.17 You may be disturbed by one odd fact. "If a laboratory observer sees the rocketship clock running slow", you reason, "then the rocket observer ought to see the laboratory clock running faster. And if this is so, then we should be able to tell which one of them is moving simply by comparing their clocks." It is true that, if there were some sort of asymmetry here, we could deduce which one was moving, in contradiction to the postulates of special relativity. But the point is that there is no asymmetry. If you repeat the entire

[7] The analysis of this experiment is complicated by the fact that one must take into account the rate difference of the two sets of clocks due to their difference in **height**; this latter effect must be calculated using Einstein's **general theory of relativity**.

argument above, only this time putting yourself on the rocketship and the light clock on a table in the laboratory, you would observe the lab clocks to be running slow, compared to your own, and by exactly the same gamma factor as before! The lab observer and the rocket observer both observe the other person's clock to be running slow. The situation is completely symmetric, as required by the principles of relativity.

How is this possible? Is there not some sort of inconsistency here? The answer is no, there is no inconsistency. But to convince you of that fact, we must go even deeper into the land of special relativity.

4.18 Before we do this, however, we must clear up one final common misconception. Students who learn relativity for the first time sometimes think that the phenomena discussed here are somehow artifacts of the observation process itself. That is, they think that the problems lie in the fact that light takes a finite time to propagate from place to place, and that this is what is responsible for the odd things such as time dilation. But a few moments thought ought to convince you that this is untrue. We never once mentioned the finite light travel time in our analysis of the light clock experiment above; we assumed the observations were instantaneous. Indeed, if you do take into account the finite travel time of light in the observation process, the above results are modified accordingly.

Length Contraction

4.19 We now discuss another odd consequence of the relativity postulates, namely the phenomenon of **length contraction**. Heuristically, we could reason as follows. If two observers are in relative motion, they each measure the same relative speed between them. But if time is different for each of them, distances must also be different, because speed is distance divided by time.

4.20 More precisely, consider the following experiment. A meter stick is at rest in the laboratory, and a rocket comes flying by at speed β:

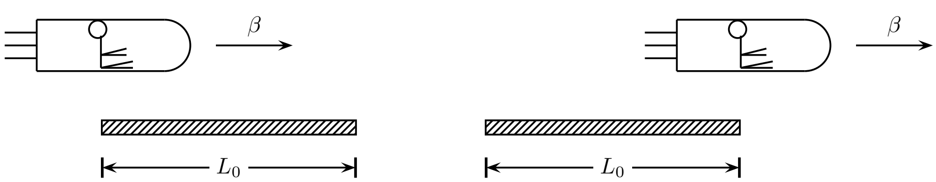


Figure 9. A rocket flies by a meter stick.

The time it takes for the rocket to pass by the entire length L_0 of the stick is, according to the laboratory observer,

$$\Delta t = \frac{L_0}{v} \tag{4.10}$$

Now, the rocket observer is timing the same two events (rocket passes by one end of the meter stick, then rocket passes by the other end of the meter stick). If the laboratory observer looks at the rocket observer's stopwatch, it reads

$$\Delta t' = \frac{\Delta t}{\gamma} = \frac{L_0}{\gamma v} \tag{4.11}$$

which is simply the result of Equation (4.8) applied to this problem. According to the rocket observer, though, the time interval between the two events is

$$\Delta t' = \frac{L'}{v} \tag{4.12}$$

where L' is the length of the meter stick according to the rocket observer. It follows from (4.11) and (4.12) that

$$L' = \frac{L_0}{\gamma} \tag{4.13}$$

In other words, the rocket observer measures the meter stick to be **shortened** or **contracted** by a factor of γ! This phenomenon is referred to as **Lorentz-Fitzgerald length contraction**, because it was first proposed by those physicists as an explanation for the Michelson-Morley experiment. [8]

4.21 Example. Suppose a rocket ship flies by a meter stick at a speed $\beta = 0.71$. Then $\gamma = 1.41$ and he measures the laboratory meter stick to be only

$$L' = \frac{L}{1.41} = 0.71 \text{ m}$$

[8] The famous Michelson-Morley experiment (see Footnote 2) was designed to test the hypthesis that light waves travel through a mysterious, all-pervasive medium known as the **aether**. If there were such a substance, Einstein's postulates would be mistaken, as one could measure one's velocity with respect to the aether. However, Michelson and Morley performed this experiment in 1887 (18 years before Einstein's paper on special relativity) and found no evidence for the aether. H.A. Lorentz suggested that the null result might be accounted for if one assumed that the lengths of objects were contracted by an amount given by (4.13) in the direction of their motion through the aether. The Kennedy-Thorndike experiment (performed in 1932) ruled out this explanation. It is sometimes mistakenly asserted that the Michelson-Morley experiment served as one of the inspirations for Einstein when he created his special theory of relativity, but this is wrong. According to most historical sources, even if Einstein was aware of the result of the experiment, it did not figure prominently in his thinking.

That is, to the rocket observer, the meter stick is only 71 cm long!

4.22 Once again, note that, if meter sticks appear shorter to the rocket observer, the same must be true of **all** objects. That is, according to the rocket observer **the entire universe is contracted by a factor of γ in the direction of motion of the rocket ship**. Also, by relativity, the same must be true in reverse. That is, according to the Earth observer, the rocket ship and everything on it is shortened by a factor of γ in the direction of motion.

4.23 It is clear from (4.13) that the degree of shortening depends on γ, which, in turn, depends on the speed of the rocket. Notice what happens when the rocket ship approaches the speed of light. Then $\beta \to 1$, so $\gamma \to \infty$, in which case its length along the direction of motion shrinks to zero (as seen by the laboratory observer)! This is one way to see why, according to the special theory of relativity, the speed of light is the speed limit for the universe. [9]

4.24 It is worth observing that **the phenomenon of length contraction only affects the dimensions in the direction of motion**. The transverse directions are completely unaffected by the Lorentz contraction. This must be so, as the following *gedanken* experiment (or thought experiment) demonstrates. Suppose our rocketship were shaped like a cylinder 1 meter in diameter, and suppose there were a tube, also one meter in diameter, into which the stationary rocket fits perfectly. Now let the rocket ship fly through this tube at high speed. If the laws of nature were such that objects contracted transversally at high speeds, then the laboratory observer would observe a gap between the rocket and the tube. But, according to the First Postulate of relativity, there ought to be no way of telling which observer was moving. So we could analyze the experiment from the point of view of the rocket observer. To her, the tube moves at high speed, and so should be contracted radially. But if that were true, the pilot would meet an untimely end instead of passing through the tube. This contradiction means that there can be no transversal contractions of objects. A similar argument rules out transversal expansions. Hence the transverse dimensions of objects must be the same for all observers.

[9] Actually, this only rules out an object approaching the speed of light from *below*, that is, from a speed less than that of light. It is theoretically possible for objects to exist that perpetually travel *faster* than the speed of light; they simply cannot ever *slow down* to the speed of light. These hypothetical particles, known as **tachyons** (from the Greek, $\tau\alpha\chi\eta\sigma$, meaning 'swift'), would possess very odd properties were they to exist. For example, they would have to travel backwards in time.

5. Proper Time and The Invariant Interval

5.1 We have now met a few of the strange phenomena that occur when objects move with relative speeds approaching the speed of light, including time dilation and length contraction. These phenomena allow for endless arguments between, say, the rocket observer and the lab observer, over whose time is really the "right" time. Of course, as we learned above, there really is no conflict, as the flow of time depends on your relative state of motion. But you might ask, "Is there anything about which the two observers can agree?" The answer is yes, and it is called the **proper time** (or, equivalently, the **invariant interval**) between two different events.

5.2 Before we can introduce the idea of proper time, let us recall a few elementary facts from Euclidean geometry. Suppose I wish to measure the distance Δs between two points A and B. One way to do this is to employ coordinates. Let (x_A, y_A) and (x_B, y_B) be the coordinates of the points A and B, respectively. Then, according to Pythagorus' famous theorem, the distance between the two points is given by

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2} \tag{5.1}$$

where

$$\Delta x = x_B - x_A \qquad \text{and} \qquad \Delta y = y_B - y_A \tag{5.2}$$

are the coordinate differences between the two points A and B. For example, if $\Delta x = 3$ m and $\Delta y = 4$ m, then $\Delta s = 5$ m:

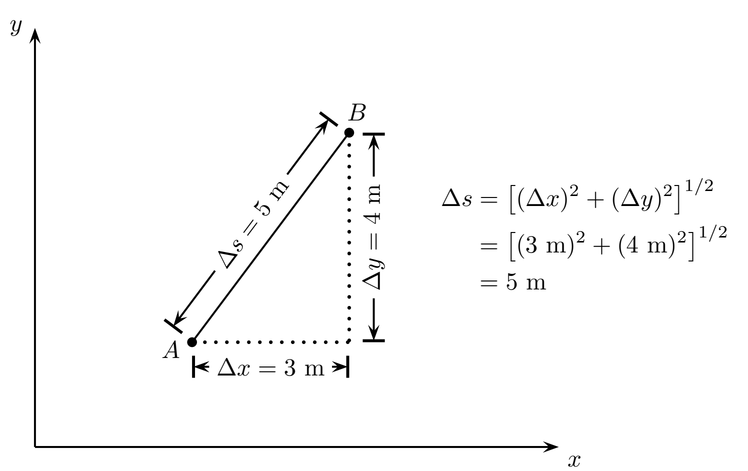


Figure 10. Euclidean geometry.

The distance (5.1) is called the **Euclidean distance** between the two points A and B.

5.3 Another way to measure the distance between the two points would be to use a different coordinate system, one that is related to the first by, say, a translation followed by a rotation through some angle. Let us call this new coordinate system the primed system. Then, in this new coordinate system the point A is represented by the pair (x'_A, y'_A) and the point B by the pair (x'_B, y'_B). Of course, in general,

$$(x_A, y_A) \neq (x'_A, y'_A)$$

and

$$(x_B, y_B) \neq (x'_B, y'_B)$$

That is, the two points A and B (which are fixed) are described by different numbers in the two different coordinate systems. But, in this new coordinate system the distance between two points is still given by the Pythagorean theorem, and, moreover, this distance is **invariant**—it is the same for both observers. This is illustrated in Figure 11:

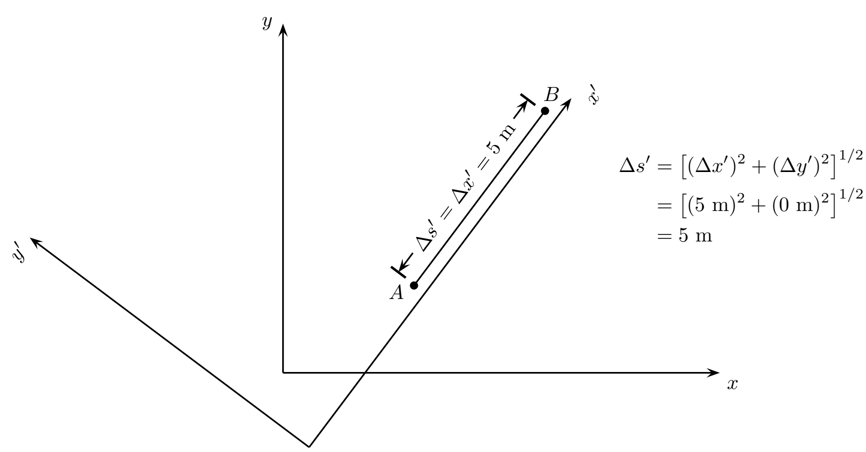


Figure 11. A translated and rotated coordinate system.

5.4 The point of the demonstration is that, although the two observers, unprimed and primed, disagree about the x and y coordinates of the two points A and B, they always compute the same distance between them:

$$\boxed{\Delta s = \Delta s'} \tag{5.3}$$

Put another way, **the distance between two points is invariant under translation or rotation**.

5.5 In special relativity, the analogue of the **(invariant) distance** between two **points** A and B in **space** is the **(invariant) interval** between two **events** A and B in **spacetime**. So we suppose now that A and B represent two events in spacetime, with coordinates (ct_A, x_A) and (ct_B, x_B) as shown

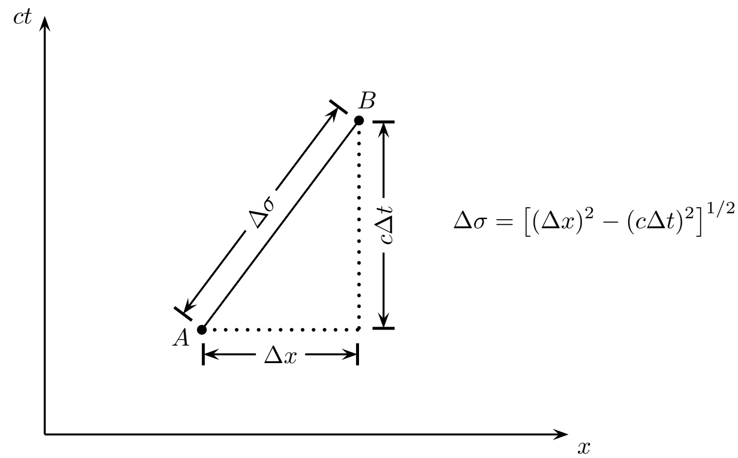


Figure 12. The (invariant) interval between two events A and B in spacetime.

Then we may define the **interval** $\Delta\sigma$ between the two events A and B to be

$$\Delta\sigma := \sqrt{(\Delta x)^2 - c^2(\Delta t)^2} \tag{5.4}$$

Right away we notice two things. First, it looks just like Pythagorus' Theorem, except for the minus sign. We will have more to say about this sign difference later, but it is worth pointing out that it is precisely this difference that distinguishes time from space. Second, we need to multiply by various factors of c in order to ensure that all the terms have units of distance.

5.6 Related to the interval $\Delta\sigma$ is the **proper time** between the two events A and B, defined as

$$\Delta\tau := \sqrt{(\Delta t)^2 - (\Delta x/c)^2} \tag{5.5}$$

The relation between them is $(\Delta\sigma)^2 = -c^2(\Delta\tau)^2$ or

$$\Delta\sigma = ic\Delta\tau \tag{5.6}$$

where $i = \sqrt{-1}$.

5.7 Notice that, when the interval between two events is real, the proper time between them is complex, and *vice versa*. We therefore distinguish three separate cases. If the proper time between two events is real, the two events are said to be **timelike separated**. If the interval between the two events is real, the events are said to be **spacelike separated**. Finally, if the proper time (and hence the interval) between the two events **vanishes**, we say the two events are **lightlike separated**. [10]

5.8 The point of these definitions is that, whereas any two observers will disagree on the times and locations of individual events, they will **always** agree on the interval (and hence the proper time) between two different events in spacetime. That is, **the interval (proper time) is the same for all observers moving relative to one another with constant speed**:

$$\boxed{\Delta\sigma = \Delta\sigma' \qquad \text{and} \qquad \Delta\tau = \Delta\tau'} \tag{5.7}$$

Indeed, the invariance of the interval (5.7) characterizes the **geometry** of spacetime. (We shall have much more to say about this later.) [11]

5.9 Before we investigate this idea further, let us first illustrate the invariance of proper time using our example of the light clock. For our example, the laboratory observer computes the proper time between the emission and receipt of the light signal to be

$$\begin{aligned} \Delta\tau &= \sqrt{(\Delta t)^2 - (\Delta x/c)^2} \\ &= \left[(4/c^2)(D^2 + (L/2)^2) - (L/c)^2\right]^{1/2} \\ &= \frac{2D}{c} \end{aligned}$$

while the rocket observer computes the proper time to be

$$\begin{aligned} \Delta\tau' &= \sqrt{(\Delta t')^2 - (\Delta x'/c)^2} \\ &= \sqrt{(2D/c)^2 - 0} \\ &= \frac{2D}{c} \end{aligned}$$

[10] The terminology just introduced is intimately connected with the **causal structure** of the universe. The only way an event A can affect an event B is if some influence propagates from A to B. If events A and B may be connected by a light signal, then A and B are lightlike separated ($\Delta\tau = 0$). If a physical object could travel from event A to event B, then A and B are timelike separated. Hence, A could causally affect B only if the two events are timelike or lightlike separated. Contrariwise, if A and B are spacelike separated, B cannot be caused by A (and *vice versa*).

[11] In what follows we speak either of the invariant interval or the proper time between two events. Statements about the invariance of the proper time are equivalent to statements about the invariance of the interval, as is evident from Equation (5.6).

They agree, as advertised.

5.10 Notice that **proper time equals coordinate time for the observer for which the two events coincide in space**. In our running example, the two events, emission and absorption of the light signal, occur at the same point in space for the rocket observer. Hence the time he measures between the two events is also the proper time between those events, as you can see from the definition (5.5). Of course, we could repeat the same experiment using the rocket observer's wrist watch or heartbeat instead of the light clock. It follows that every observer measures time's passage according to proper time. This is why proper time is called "proper".

6. Lorentz Transformations I

6.1 We now proceed to find the general form of the transformation relating the time and space coordinates of two inertial reference frames. This was first done by H.A. Lorentz, so the resulting transformations are called **Lorentz transformations**, in his honor. [12] To begin, recall what sort of transformation we would expect according to Newtonian physics. Suppose the rocket ship were moving with speed v to the right. Then, according to the **Principle of Galilean Relativity**, the coordinate systems of the two reference frames are related by the formula

$$\boxed{\begin{aligned} t' &= t \\ x' &= x - vt \end{aligned}} \tag{6.1}$$

Note that time is the same in the two frames; also, according to the rocket observer, all stationary objects in the lab frame move to the left with speed v. The Principle of Galilean relativity is so simple and obvious, it seems hard to believe that it is wrong.

6.2 To discover the correct form of the transformation relating the coordinates of two inertial reference frames, we first demand on physical grounds that the transformation be

[12] In three dimensions, the two frames may be translated and rotated relative to one another. So we may simplify matters by assuming that we have already performed the necessary rotation to line up their axes and translation to bring their origins into coincidence. Of course, we are temporarily restricting ourselves to one dimension, so the only thing a rotation can do is reflect the positive x direction about the origin. We therefore assume, as above, that we have performed such a reflection in order to ensure that both observers agree on the positive x direction. It is then easy to translate one of the coordinate systems to ensure that the origins coincide.

linear. [13] For example, the transformation (6.1) is linear. The reason for this demand is simple: **inertial observers move along straight lines in all inertial reference frames**. To maintain this condition, a straight line in one frame must be mapped to a straight line in another. The only transformation that accomplishes this is a linear one.

6.3 The most general linear transformation between the coordinates of the two reference frames is

$$\begin{aligned} t' &= At + Bx \\ x' &= Ct + Dx \end{aligned} \tag{6.2}$$

Because this transformation is linear, it follows that

$$\begin{aligned} \Delta t' &= A\Delta t + B\Delta x \\ \Delta x' &= C\Delta t + D\Delta x \end{aligned} \tag{6.3}$$

where $\Delta x = x_2 - x_1$, for example, denotes the coordinate difference of two distinct events in the lab frame, and so forth.

6.4 Now we use the results of our previous analysis, namely time dilation and length contraction. We agreed that the rocket observer sees the lab clocks running slow, and *vice versa*. For example, if we let the two events in the laboratory be (1) a clock registers t_1, and (2) the same clock registers t_2, then for these two events we have $\Delta x = 0$ (as they take place at the same location in the laboratory frame), and $\Delta t' = \gamma\Delta t$ (the laboratory clock runs slow, according to the rocket observer). It therefore follows from (6.3) that $A = \gamma$.

6.5 On the other hand, suppose a meter stick comes flying by the lab at speed β. We agreed that the laboratory observer measures it to be contracted in length. We define two events: (1) lab observer marks the position of the left end of stick as it comes flying by, and (2) lab observer marks the position of the right end of the stick as it comes flying by. Clearly, if the lab observer wants to measure the length of the stick, he must make these measurements **simultaneously**. Hence $\Delta t = 0$. But the stick is Lorentz contracted, so $\Delta x' = \gamma\Delta x$. Comparing these data with Equation (6.3) we see that $D = \gamma$.

6.6 Next, consider another experiment in which the laboratory observer measures the position of the nose of the rocket at two different times. (You could imagine that the nose of the rocket goes through two photogates in the laboratory.) These two events take place

[13] For more on linearity, see Appendix A.

at the same location in the rocket frame, because, relative to the rocket observer, the nose of the rocket is always a fixed distance away from him. Hence $\Delta x' = 0$. According to (6.3), then,

$$0 = C\Delta t + D\Delta x \Longrightarrow \frac{C}{D} = -\frac{\Delta x}{\Delta t}$$

But $\Delta x/\Delta t$ is just v, the velocity of the rocket, so $C/D = -v$. Combining the results of this Section with those of Section **6.5**, we find that $C = -v\gamma$.

6.7 It remains to find the last coefficient B in (6.3). This can be done in many ways, but the most straightforward is to use the invariance of the interval (5.7). We have

$$\begin{aligned}(\Delta x')^2 - c^2(\Delta t')^2 &= (-v\gamma\Delta t + \gamma\Delta x)^2 - c^2(\gamma\Delta t + B\Delta x)^2 \\ &= -\gamma^2(c^2 - v^2)(\Delta t)^2 + 2(\Delta x)(\Delta t)(-v\gamma^2 - Bc^2\gamma) + (\gamma^2 - c^2B^2)(\Delta x)^2\end{aligned}$$

This must equal $(\Delta x)^2 - c^2(\Delta t)^2$, so looking at the cross term we see immediately that $B = -v\gamma/c^2 = -\beta\gamma/c$.

6.8 Putting all these results together, we have

$$\boxed{\begin{aligned} c\Delta t' &= \gamma c\Delta t - \beta\gamma\Delta x \\ \Delta x' &= -\beta\gamma c\Delta t + \gamma\Delta x \end{aligned}} \tag{6.4}$$

This transformation of coordinates between inertial frames in relative motion is called a **Lorentz boost**. Written out without abbreviations it reads (choose one event to be the origin):

$$\begin{aligned} t' &= \frac{t - \frac{vx}{c^2}}{\sqrt{1 - (\frac{v}{c})^2}} \\ x' &= \frac{x - vt}{\sqrt{1 - (\frac{v}{c})^2}} \end{aligned} \tag{6.5}$$

Comparing this to the (incorrect) Galilean transformation (6.1) we see that the two main differences are (i) time itself gets transformed between the two frames, and (ii) distances get transformed with an additional factor of γ. Of course, (6.5) reduces to (6.1) when the rocket speed is small compared to light, as it must.

6.9 EXERCISE. Show that the inverse Lorentz boost is given by

$$\boxed{\begin{aligned} c\Delta t &= \gamma c\Delta t' + \beta\gamma\Delta x' \\ \Delta x &= \beta\gamma c\Delta t' + \gamma\Delta x' \end{aligned}} \tag{6.6}$$

That is, invert the Equations (6.4) to find the transformation that takes us from the primed coordinates to the unprimed coordinates. Explain why you could have done this without any computations at all.

6.10 In reality, we live in a world which has three spatial dimensions, not one, as we have been assuming. In three dimensions, the Lorentz boost (6.4) must be supplemented by the conditions

$$\begin{aligned} \Delta y' &= \Delta y \\ \Delta z' &= \Delta z \end{aligned} \tag{6.7}$$

assuming the rocket is moving in the $+x$ direction. In other words, the y and z directions are completely unaffected, as we argued in Section **4.24**.

6.11 Equations (6.4) and (6.6) are entirely equivalent; the particular set we choose to employ is solely a matter of convenience.

7. Applications of the Lorentz Transformation

We now turn to some applications of the formalism we have taken great pains to develop. In the process we will gain a deeper insight into the implications of Einstein's theory. We will be particularly interested in various paradoxes that seem to arise, and their satisfactory resolution within the framework of special relativity.

Time Dilation Revisited

7.1 We begin by looking at a situation similar to that of the light clock that started all this. Suppose a rocket goes by the lab at a speed of $\beta = 0.8$. On the rocket there is a ticking clock. The astronaut observes the clock register 00:00 and then 01:00—that is, he observes his clock for one minute. How long does the Earth observer measure for this process? This is precisely the sort of question the Lorentz transformation is designed to answer.

For these two events we have

$$\Delta x' = x'_2 - x'_1 = 0$$

because they take place at the same location in the rocket frame, and

$$\Delta t' = t'_2 - t'_1 = 1 \text{ minute}$$

According to the Lorentz transformation Equations (6.6)

$$\Delta t = \gamma(\Delta t') + \beta\gamma\left(\frac{\Delta x'}{c}\right) = \gamma\Delta t'$$

In our case

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} = \frac{1}{\sqrt{1-(0.8)^2}} = \frac{1}{\sqrt{1-0.64}} = \frac{1}{\sqrt{0.36}} = \frac{1}{0.6} = \frac{10}{6}$$

Hence

$$\Delta t = \left(\frac{10}{6}\right)(1 \text{ min}) = 1 \text{ min } 40 \text{ sec}$$

That is, the Earth observer observes 1:40 elapse on his clock while the rocket observer measures only 1:00. As expected, the Earth observer sees the rocket clock running slow.

7.2 So far, so good. Now, suppose we reverse the situation. That is, we put the clock in the laboratory and have the astronaut watch the Earth clock. How much time passes according to the rocket observer when one minute elapses on the Earth clock? The **wrong** way to proceed is as follows. You might be tempted to say that, because, **in the previous case**, 100 seconds of Earth time passes for 60 seconds of rocket time, **in the case at hand**, 60 seconds of Earth time passes for $60 \times (60/100) = 36$ seconds of rocket time. If this were true, then both the Earth observer and the rocket observer would agree that the rocket clock was running slower than the lab clocks, in which case they could both agree that it was the rocket that was moving, and not the Earth, in contradiction with the postulates of relativity.

7.3 The correct solution of the problem requires that we be much more precise about the two events. In the first experiment, the clock was at rest **in the rocket ship**. In the case at hand, the clock is at rest **in the laboratory**. This makes all the difference in the world. For now, the two events are (1) Earth clock registers 00:00, (2) Earth clock registers 01:00. For these two events we have

$$\Delta x = x_2 - x_1 = 0$$

and **not** $\Delta x' = 0$, as was the case previously. Similarly, we now have

$$\Delta t = t_2 - t_1 = 1 \text{ min}$$

So, using Equations (6.4), we find

$$\Delta t' = \gamma\Delta t - \beta\gamma(\Delta x/c) = \gamma\Delta t = \frac{10}{6} \text{ min}$$

as we should. Both observers agree that the other person's clock is running slow.

7.4 Although we have succeeded in reconciling the two observers' points of view, you may still be left with an uneasy feeling. Sure, the equations are consistent, but physically how is it possible? It is very difficult to respond to such feelings, as their assuagement lies more in the realm of psychology than physics. This is simply the way things are. You should at least try to convince yourself that there is no paradox, *i.e.,* that the two experiments are perfectly consistent. We will have another opportunity to do this later when we consider the twin paradox.

Length Contraction Revisited

7.5 Suppose a rocket flies by the laboratory with a speed $\beta = 0.8$. On the rocket there is a meter stick oriented along the direction of motion. How long is the meter stick according to the Earth observer? As always, it behooves us to be very careful about the two events we are comparing. How does one measure the length of anything? Well, you put your ruler next to the object and mark off the two ends of the object on your ruler. If the object is moving, you must be careful to mark the two ends **at the same time**, otherwise you will not measure the true length of the object.

7.6 So, to be entirely unambiguous, we carry out the following experiment. You are in the laboratory, and as the meter stick flies by in the rocket, you take two meat cleavers and simultaneously strike your ruler, with one cleaver at the back of the flying meter stick and one at the front. Then, at your leisure, you observe the distance between the two gashes in your ruler. The two events are (1) meat cleaver strikes trailing edge of flying meter stick and (2) meat cleaver strikes leading edge of flying meter stick. As we have taken great pains to emphasize, for these two events we have $\Delta t = 0$. We wish to find Δx. We know $\Delta x' = 1$ meter. So, using Equations (6.4) we find

$$\Delta x' = -\beta\gamma c\Delta t + \gamma\Delta x \Longrightarrow \Delta x = \Delta x'/\gamma = 0.6\Delta x' = 0.6 \text{ meter}$$

As expected, you measure the meter stick to be shorter than a meter; specifically, you measure its length to be 60 cm.

7.7 Of course, if we perform the experiment in reverse we would find that the rocket observer measures the Earth meter sticks to be 60 cm in length. As before, there is no paradox, but it is not exactly clear why. To bring out this point more clearly, we now consider the famous pole and barn paradox.

8. The Pole and Barn Paradox

8.1 Farmer John is sittin' in his corn field one day when all of a sudden this pole vaulter appears out of nowhere. The pole vaulter is carrying a pole 20 m long and is running at a large fraction of the speed of light right toward Farmer John's barn, which is only 10 m long. Both doors of the barn are open, and the pole vaulter (as he cannot easily stop) aims to run right through the barn.

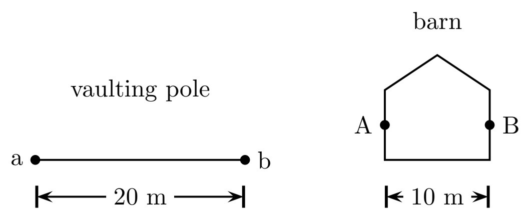


Figure 13. Pole and barn both at rest in the lab frame.

Now, Farmer John is a mite upset that this city slicking pole vaulter is trampling his corn, so he decides to slam the doors on the pole vaulter and trap him in the barn. The problem is, the barn is only 10 m long. But the pole vaulter is moving so fast (he is wearing good running shoes) that his pole is Lorentz contracted to 10 m in the rest frame of the barn. So, Farmer John reasons, there will be no problem trapping the pole vaulter in the barn.

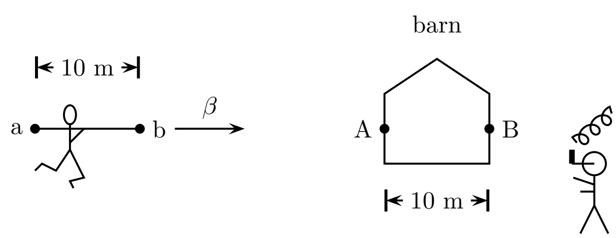


Figure 14. The situation as seen by Farmer John.

8.2 This would be the end of it, except for one small problem. According to the pole vaulter, his pole is 20 m long. The barn, on the other hand, is moving toward him at such a large fraction of the speed of light, that its length is Lorentz contracted from 10 m to 5 m! So the pole vaulter, figuring out what Farmer John is up to, shouts to him not to slam the front door of the barn lest he sever the pole.

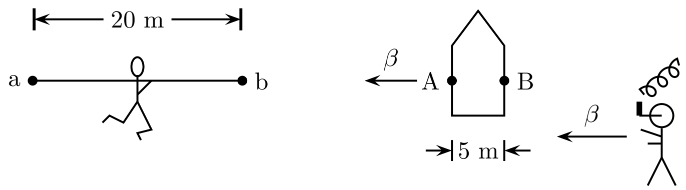


Figure 15. The situation as seen by the pole vaulter.

8.3 The paradox, of course, is "who is right?" Surely, either the pole will fit completely inside the barn or it will not. All we have to do is perform the experiment. When we do, we will discover either that the pole fits inside the barn or it does not. So either Farmer John or the pole vaulter must be wrong, and hence the theory of special relativity must be wrong.

8.4 The resolution of the paradox is that **both the farmer and pole vaulter are right**! According to Farmer John, there is an instant during which the pole fits completely in the barn (before it bursts through the back of the barn). According to the pole vaulter, the pole **never** fits entirely within the barn, but rather bursts through the back door of the barn before the front door closes (although the pole never gets severed—that would be a paradox!).

8.5 To see exactly what happens, let us again be very precise about the two events in question. We label as (1) the event: the trailing edge a of the pole passes the front door A of the barn, and as (2) the event: the leading edge b of the pole passes the back door B of the barn. Before we can use the Lorentz boost formula, we must determine the speed of the pole vaulter. But this is easy. According to Farmer John, the pole is Lorentz contracted from 20 m to 10 m. From Equation (4.13) applied to this problem we have

$$L = \frac{L_0}{\gamma} \Longrightarrow \gamma = \frac{L_0}{L} = \frac{20\text{ m}}{10\text{ m}} = 2$$

Now we solve for β:

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

$$\Longrightarrow \beta = \sqrt{1-\frac{1}{\gamma^2}} = \sqrt{1-\frac{1}{4}} = \sqrt{\frac{3}{4}} = 0.87$$

8.6 Based upon our choice for the two events, we have

$$\Delta t = t_2 - t_1 = 0$$

and

$$\Delta x = x_2 - x_1 = 10 \text{ m}$$

From the Lorentz boost equations (6.4) we have

$$c\Delta t' = \gamma c\Delta t - \beta\gamma\Delta x$$
$$\Delta x' = -\beta\gamma c\Delta t + \gamma\Delta x$$

The second of these yields

$$\Delta x' = \gamma\Delta x = 2(10 \text{ m}) = 20 \text{ m}$$

What does this mean? It means that, according to the pole vaulter, the two events, (1) front door hits trailing edge of pole and (2) back door hits leading edge of pole occur 20 m apart in space. This has to be true, for the pole is 20 m long according to the pole vaulter.

8.7 The odd thing comes next. The time between the two events, as measured by the pole vaulter, is

$$\Delta t' = -\beta\gamma\Delta x/c = -\frac{(0.87)(2)(10 \text{ m})}{3 \times 10^8 \text{ m/s}} = -58 \text{ ns}$$

This remarkable result means that, according to the pole vaulter, the two events **do not take place at the same time**! This is called **the Relativity of Simultaneity**. Two events that are simultaneous in one reference frame are not necessarily simultaneous in another. Indeed, we have

$$\Delta t' = t'_2 - t'_1 < 0 \Longrightarrow t'_2 < t'_1$$

According to the pole vaulter, event 2 (breaking through the back door B of the barn with the leading edge b of the pole) occurs 58 nanoseconds before event 1 (the front door A slamming on the trailing edge a of the pole). In other words, even though Farmer John slams the doors at the same time in his frame, according to the pole vaulter the back door B closes first, and the front door A closes 58 ns later. In particular, this means that the pole does not get severed, because, according to the pole vaulter, most of its length is sticking out the back of the barn by the time the front door is closed. Relativity is perfectly consistent, if a bit odd.

9. Relativistic Velocity Addition Formula

9.1 As we learned, the development of the special theory of relativity was inspired in part by Einstein's musings on the speed of light as measured by an observer carried along with the light beam. Equipped with the Lorentz transformation equations, we may now provide a satisfactory resolution of this question. The question may be stated as follows: A rocket is moving with speed β_R relative to the laboratory frame. On the rocket, the astronaut tosses a ball forward with speed β' (as measured by him, of course). What is the speed β of the ball as measured by the laboratory observer?

9.2 Note that we have introduced a slight modification in the notation. Now β_R is the velocity of the rocket relative to the lab ('R' stands for 'rocket') whereas β is the speed of the baseball in the lab frame, **not** the relative speed of the two reference frames. Also, $\gamma_R = (1 - \beta_R^2)^{-1/2}$.

9.3 Suppose that we set up two photogates in the laboratory frame, and suppose that, as the rocket comes flying by, the astronaut throws the ball forward in such a way that it goes through the two photogates. Then event (1) is the ball going through the first photogate, and event (2) is the ball going through the second photogate.

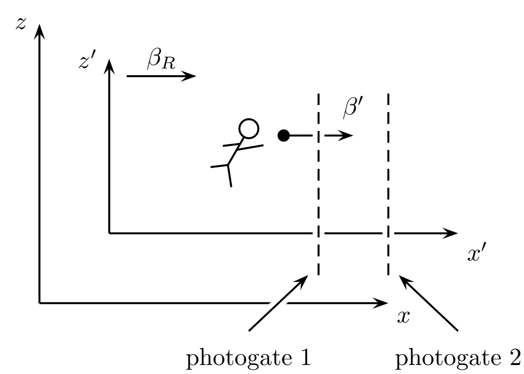


Figure 16. The astronaut throws the ball forward at speed β' relative to the rocket, while the rocket moves at speed β_R relative to the lab. Events 1 and 2 correspond to the ball going through the first and second laboratory photogates, respectively.

The speed of the ball, as measured by the laboratory observer is clearly

$$\beta = \frac{\Delta x}{c\Delta t} \tag{9.1}$$

According to the Principle of Galilean Relativity, we have (*cf.* Equation (6.1)—here we use the inverse transformation):

$$\begin{aligned} t &= t' \\ x &= x' + v_R t' \end{aligned} \tag{9.2}$$

where v_R is the speed of the rocket relative to the lab. Plugging this into (9.1) we conclude that

$$\begin{aligned} \beta &= \frac{\Delta x' + v_R \Delta t'}{c\Delta t'} \\ &= \frac{\Delta x'}{c\Delta t'} + \frac{v_R}{c} \\ &= \beta' + \beta_R \end{aligned} \tag{9.3}$$

where we used the fact that the speed of the baseball, as measured by the rocket observer, is

$$\beta' = \frac{\Delta x'}{c\Delta t'} \tag{9.4}$$

Multiplying through by c Equation (9.3) reads $v = v_R + v'$, which is just the usual nonrelativistic formula for the addition of velocities: If the rocket is moving forward at 5 mph and the astronaut tosses the ball forward at 3 mph, the speed of the ball relative to the ground is 8 mph. What could be more natural?

9.4 The point, of course, is that the formula (9.3) is wrong. More precisely, it is only true when the velocities involved are small compared to the speed of light. The correct relativistic formula is obtained by starting from equation (9.1) but employing the Lorentz transformation formula (6.6) instead. This gives (with our new notation)

$$\begin{aligned} \beta &= \frac{\beta_R \gamma_R c\Delta t' + \gamma_R \Delta x'}{\gamma_R c\Delta t' + \beta_R \gamma_R \Delta x'} \\ &= \frac{\beta_R + (\Delta x'/c\Delta t')}{1 + \beta_R(\Delta x'/c\Delta t')} \end{aligned}$$

Using (9.4) therefore yields

$$\boxed{\beta = \frac{\beta_R + \beta'}{1 + \beta_R \beta'}} \tag{9.5}$$

This is called the **relativistic velocity addition formula**.

9.5 Notice that, for small speeds, the relativistic formula (9.5) reduces to the non-relativistic formula (9.3). This is easily seen, because for small speeds both β_R and β' are much less than one. For example, suppose the rocket is moving at $\beta_R = 10^{-4}$ (*i.e.,* 30,000 m/s) and the baseball is moving at a speed $\beta' = 10^{-4}$ with respect to the rocket. Then the non-relativistic formula (9.3) gives $\beta = 10^{-4} + 10^{-4} = 2 \times 10^{-4}$. The relativistic formula (9.5) gives $\beta = (10^{-4} + 10^{-4})/(1 + 10^{-8}) \approx 2 \times 10^{-4}$. In other words, the difference between the two formulae is immeasurably small. This is why humans did not discover the laws of special relativity earlier. But when the speeds involved are close to that of light the differences between the two formulae are quite noticeable, as the following example illustrates.

9.6 Example. Suppose the rocket is moving forward with speed $\beta_R = 0.8$, and suppose the astronaut throws the ball forwards at a speed $\beta' = 0.8$ (he has been working on his fast balls). According to the Galilean formula (9.3), the speed of the ball, as measured from the lab, should be $\beta = \beta_R + \beta' = 0.8 + 0.8 = 1.6$, which is greater than the speed of light. On the other hand, the relativistic velocity addition formula (9.5) gives $\beta = (0.8 + 0.8)/(1 + (0.8)^2) = 0.976$, which is less than the speed of light, as it must be.

9.7 Example. Similarly, suppose that $\beta_R = 0.8$ and $\beta' = 1$. That is, suppose the astronaut turns on a flashlight in his rocket, sending a light beam in the forward direction. According to the laboratory observer, the speed of light beam is $\beta = (0.8+1)/(1+0.8) = 1$! That is, the laboratory observer measures the speed of the light beam to be precisely the speed of light, in accord with Postulate 1. The Galilean formula, on the other hand, yields $\beta = 1 + 0.8 = 1.8$, which is experimentally incorrect.

10. The Twin Paradox

10.1 The pole and barn paradox brings out very clearly some of the peculiarities of length contraction in special relativity. Now we turn to some of the peculiarities of time dilation. We noted that observers in relative motion each measure the other person's clock to run slow. As we discussed, this is the perfectly consistent conclusion of the theory of special relativity. But you may have wondered whether it leads to some sort of paradox. Here is one such paradox.

10.2 On their 21$^{\text{st}}$ birthdays, Peter leaves his twin brother Paul behind on Earth and flies off at a speed of $\beta_R = 0.96$ to a star 25 light years distant. [14] Let us call the star

[14] A **light year** is the distance light travels in a year. It is a **distance** not a **time**.

Vega (even though Vega, in the constellation Lyra, is closer to 23 light years away). Upon arriving at Vega he turns right around and returns to Earth at the same speed. We may ask: how old is each twin upon Peter's return?

10.3 Even without calculating anything, we know how old Paul will be. Peter is traveling so close to the speed of light that he essentially covers one light year in one year. So if it takes him 25 years out plus 25 back, when he returns Paul must be $21+50=71$ years old.

10.4 To find out how old Peter is, we measure the time for the journey according to Peter's chronometer. As the trip is clearly symmetric, we need only compute the time as measured by Peter for the outward leg, then multiply this by two. We may therefore take the two events to be: (1) Peter leaves the Earth, and (2) Peter arrives at Vega. According to Peter, the distance from Earth to Vega is Lorentz contracted. The gamma factor corresponding to a speed of $\beta_R = 0.96$ is $\gamma_R = 3.57$, so according to Peter the distance from Earth to Vega is only

$$\Delta x' = \Delta x/\gamma_R = (25 \text{ lightyears})/3.57 = 7 \text{ lightyears}$$

As he is traveling almost at the speed of light, the trip takes approximately 7 years. Hence, according to Peter, the total trip took only 14 years; when he returns to Earth, he is $21+14=35$ years old.

10.5 So far, this is unusual, but not paradoxical. Paul stayed at home and reached the age of 71 while Peter traveled to Vega and back and arrived home a young 35 years old. The paradox arises because, according to special relativity, it ought to be impossible to tell who actually did the traveling. That is, everything is relative, so we could have looked upon the trip as if Peter stayed in his rocket ship at one location in space while the entire Earth (carrying Paul with it) sped off in the opposite direction and then returned. In that case, Paul would have aged 14 years while Peter aged 50 years, and so when they again meet, their ages should be the reverse of the above! But we could imagine doing this experiment and comparing their ages. So who is younger after the trip, Peter or Paul? Or are they somehow the same age? Does this mean that special relativity is fundamentally flawed?

10.6 Of course, there is no paradox. But the reason lies outside the domain of validity of special relativity. The asymmetry in the situation is clearly that **one of the twins must accelerate**, namely Peter. He must accelerate at the beginning of his journey, slow down, stop, reverse, accelerate back toward Earth, and then slow down again to come home. It is the acceleration that is responsible for the fact that one of the twins is older and one of them is younger when they meet again. But the theory of special relativity only applies to **inertial observers**, and when Peter accelerates he is certainly no longer inertial. The theory that accounts for the physics of non-inertial observers is Einstein's **General Theory of Relativity**.

10.7 Does that mean we cannot resolve the twin paradox within the province of special relativity? No, we may do so, provided we are much more careful. [15] The key to incorporating accelerations into special relativity lies with the idea of the **instantaneously comoving reference frame**. This is an **inertial** reference frame that matches velocities with an accelerating frame for just an instant. We may understand the physics of accelerating frames within special relativity by viewing everything from the point of view of the entire (infinite) family of instantaneously comoving reference frames that parallel the trajectory of the accelerating frame.

10.8 We proceed as follows. First, note that, in the original twin paradox, Peter moves at constant velocity on the outbound leg of the trip, then instantaneously switches directions at Vega. Clearly, the problem is not changed significantly if we allow Peter to smoothly decelerate at Vega and then accelerate back toward the Earth. In fact, we may suppose that Peter takes a luxury space liner that advertises "One Earth Gravity All the Way". This way he does not have to endure acclerations that would turn him into jelly. So his flight plan calls for an acceleration of one g away from Earth, turn around at the half-way point between Earth and Vega, deceleration at one g for the second quarter of the trip until he stops at Vega, then reversing the process on the return leg. If we plot Peter's worldline (*cf.* Section **14.2**), it looks something like Figure 17.

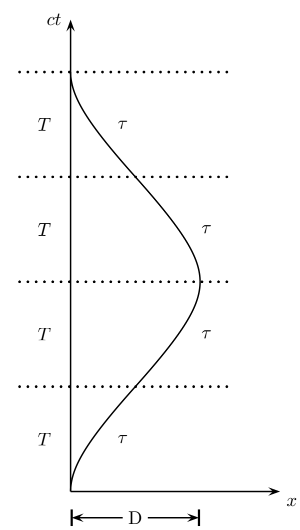


Figure 17. Peter's worldline (as seen by Paul) for his trip to Vega and back.

In the figure, D represents the distance from Earth to Vega, and we have broken the journey up into quarters. According to Paul's clock on Earth, each quarter of the journey

[15] The remainder of Chapter **10** may be skipped on a first reading.

takes a time T, while according to Peter's clock, each quarter takes a (proper) time τ. Our mission, should we choose to accept it, is to find the precise relationship between T and τ.

10.9 We will need an expression for the transformation of accelerations between reference frames. That is, Peter records an acceleration of one Earth gravity **in his reference frame**. But this cannot be the acceleration as measured by Paul, because if Peter maintained this acceleration forever he would eventually exceed the speed of light, which is impossible.

10.10 To find the relation between the two accelerations, we recall the relativistic addition formula (9.5) for velocities

$$\beta = \frac{\beta_R + \beta'}{1 + \beta_R \beta'}$$

where β_R was the speed of the rocket relative to the lab, β' was the speed of the baseball relative to the rocket, and β was the speed of the baseball relative to the lab. We want to consider the case in which the baseball has a jet pack attached to it, so that it accelerates relative to the rocket (and hence relative to the lab). In that case $\beta' = \beta'(t)$ (that is, the speed of the baseball changes with time relative to the rocket) so $\beta = \beta(t)$, but β_R, the speed of the rocket, is time independent. We wish to compute a, the acceleration of the baseball as seen in the lab, in terms of a', the acceleration of the baseball relative to the rocket.

10.11 At this point we must pause to discuss **units**. The actual speed of the baseball relative to the lab is $c\beta$, so the acceleration relative to the lab is $d(c\beta)/dt = c(d\beta/dt)$. But the factors of c are just annoying, so we agree to drop them all and put them back in at the end. We may do this by choosing units so that distances are all measured in units of **light-seconds**. Then the speed of light would be 1, because light travels one light-second every second. So in all our formulae we may set $c = 1$, provided we remember to use the right units. At the end of the day we may convert back to normal units by restoring the missing factors of c.

10.12 With this choice of units, we may write

$$
\begin{aligned}
a = \frac{d\beta}{dt} &= d\left[\frac{\beta_R + \beta'}{1+\beta_R\beta'}\right] \div dt \\
&= \frac{d\beta'(1+\beta_R\beta') - (\beta_R+\beta')\beta_R\, d\beta'}{(1+\beta_R\beta')^2} \div (\gamma_R\, dt' + \beta_R\gamma_R\, dx') \\
&= \frac{1}{(1+\beta_R\beta')^2}\left[\frac{d\beta' + \beta_R\beta'\, d\beta' - \beta_R^2\, d\beta' - \beta_R\beta'\, d\beta'}{\gamma_R\, dt' + \beta_R\gamma_R\, dx'}\right] \\
&= \frac{1}{(1+\beta_R\beta')^2}\left[\frac{(d\beta'/dt')(1-\beta_R^2)}{\gamma_R + \beta_R\gamma_R(dx'/dt')}\right] \\
&= \frac{1}{(1+\beta_R\beta')^2}\left[\frac{a'(1-\beta_R^2)^{3/2}}{1+\beta_R\beta'}\right]
\end{aligned}
$$

where we used the infinitessimal version of (6.6) and the definition of β' and γ_R. Hence we conclude

$$
\boxed{a = \frac{a'}{\gamma_R^3}(1+\beta_R\beta')^{-3}} \tag{10.1}
$$

10.13 Now we return to the twin paradox. We wish to compute the acceleration of Peter's ship according to Paul. We already said that Peter feels an acceleration of one g, but what does this mean? It means that Peter is accelerating at one g **relative to an instantaneously comoving reference frame**. We may use the results of the above analysis by interpreting β as Peter's velocity as measured by Paul, β_R as the velocity of the instantaneously comoving frame relative to Paul, and β' as Peter's velocity relative to the instantaneously comoving frame. But the comoving frame is **comoving**, which means $\beta' = 0$. So the formula (10.1) simplifies to become

$$
a = \frac{a'}{\gamma_R^3} = \frac{g}{\gamma_R^3} \tag{10.2}
$$

10.14 Two other things happen when we switch to the instantaneously comoving frame. First, β becomes β_R, because the Peter's speed, as measured by Paul, is precisely the speed of the comoving reference frame. Second, β_R becomes time dependent, because Peter is accelerating. Hence (10.2) becomes

$$
\frac{d\beta_R}{dt} = g(1-\beta_R^2)^{3/2}
$$

This is a differential equation for β_R, which we can solve. After a little manipulation we get

$$\frac{d\beta_R}{(1-\beta_R^2)^{3/2}} = g\,dt$$

Integrating both sides yields

$$\frac{\beta_R}{\sqrt{1-\beta_R^2}} = gt + c$$

We choose the constant c of integration to be zero because at $t = 0$ the rocket is not moving. Solving for β_R as a function of t then gives

$$\beta_R = \frac{gt}{\sqrt{1+(gt)^2}} \tag{10.3}$$

Notice that, as expected, β_R approaches 1 as t approaches infinity (*i.e.,* constant acceleration in rocket frame causes the rocket to approach the speed of light in the lab frame).

10.15 Integrating (10.3) again we get

$$x_R = \int \frac{gt\,dt}{\sqrt{1+(gt)^2}}$$

or

$$x_R = \frac{1}{g}\left[\sqrt{1+(gt)^2} - 1\right] \tag{10.4}$$

where we have chosen the integration constant so that $x = 0$ at $t = 0$.

10.16 Finally, we are in a position to determine how spaceship time (proper time) compares to Earth time for the first quarter of the journey (see Equation 14.10):

$$\begin{aligned}\Delta\tau &= \int d\tau = \int_0^T \frac{d\tau}{dt}\,dt = \int_0^T \sqrt{1-\beta_R^2}\,dt \\ &= \int_0^T \frac{dt}{\sqrt{1+(gT)^2}} \\ &= \frac{1}{g}\ln\left\{gT + \sqrt{1+(gT)^2}\right\}\end{aligned} \tag{10.5}$$

Also, we may compute the maximum displacement from the Earth that the spacecraft reaches (by doubling its travel distance in time T as measured on Earth):

$$x_{\max} = \frac{2}{g}\left\{\sqrt{1+(gT)^2} - 1\right\} \tag{10.6}$$

and the maximum speed it reaches on its journey (at the quarter distance mark)

$$\beta_{\max} = \frac{gT}{\sqrt{1+(gT)^2}} \tag{10.7}$$

10.17 Restoring all the missing powers of c, Equations (10.5), (10.6), and (10.7) become

$$\Delta\tau = \frac{c}{g}\ln\left\{\frac{gT}{c} + \sqrt{1+(gT/c)^2}\right\} \tag{10.8}$$

and

$$x_{\max} = \frac{2c^2}{g}\left\{\sqrt{1+(gT/c)^2} - 1\right\} \tag{10.9}$$

and

$$\beta_{\max} = \frac{gT}{c}\sqrt{1+(gT/c)^2} \tag{10.10}$$

respectively.

10.18 Collecting all these results together, we obtain the following table: [16]

Earth Time ($4T$)	Spaceship Time ($4\Delta\tau$)	$x_{\max}$	$\beta_{\max}$
4 days	4 days$-\frac{1}{2}$ sec	73×10^6 km (Mars?)	0.0028
3 months	3 months$-$91 min	35 light hours	0.065
4 years	3.5 years	0.85 lyrs	0.72
11.7 yrs	7.1 yrs	4.2 lyr (α-Centauri)	0.95
48.5 yrs	12.5 yrs	22.4 lyr	0.9968
57.0 yrs	13.1 yrs	26.6 lyr	0.9977
2350 yrs	27.5 yrs	1170 lyr	0.9999986
4×10^6 yrs	56.4 yrs	2×10^6 lyr (Andromeda)	$1 - 5 \times 10^{-13}$

The meaning of the table is as follows. An astronaut who travels at a constant acceleration of 1 Earth gravity (except for a very brief time when he is weightless as the ship turns around at the two half-way points) ages considerably less than an Earth based observer the farther he travels. This is because the farther he travels at a constant 1 g, the longer he spends at high velocities relative to the Earth.

10.19 It is worth pointing out a few features of the data. If the astronaut goes to Mars, he comes back half a second younger than his stay at home twin. If he goes as far as the nearest star (although actually Proxima Centauri is a little closer than its twin) he comes

[16] Adapted from L. Marder, *Time and the Space Traveler* (University of Pennsylvania Press, Philadelphia, 1974).

back over four and a half years younger than his twin. Finally, the farthest distance that the astronaut can travel in his lifetime is about 2 million light years, which would carry him as far as the **nearest** galaxy, Andromeda. When he returns, having aged about 56 years, Earth has aged four million years. Indeed, it may not be here when he gets back! Considering that Andromeda is the nearest of billions of galaxies, these data suggest that, as long as we are subject to the rules of special relativity, we will never be able to explore anything but the closest stars in our universe. Of course, we can create space arks and have our descendents visit other stars, but this is somewhat unsatisfying. Fortunately, there may be a way to use **general relativity** to create a **warp drive** and circumvent special relativistic limits, but this must await another day.

11. Lorentz Transformations II

11.1 In Chapter **6** we discovered the form of the transformation relating two inertial reference frames in relative motion. But, although each step was justified, the overall result is still a little mysterious. In this Chapter we develop a way of understanding the Lorentz transformation **geometrically**. In the process, we will gain a deeper insight into its meaning.

Minkowski Spacetime

11.2 Consider again two dimensional space. Until we specify that the distance between any two points is to be computed using the Euclidean distance function (5.1), two dimensional space is just a collection of points. [17] We could, if we were so inclined, put a different distance function on the space, in which case it would have a different **geometry**. A geometry on a space is something that tells you the distance between points and the angle between vectors. We call the two dimensional space with which we have been dealing $\mathbb{R}^2$ and the usual distance function the **Euclidean metric**.

11.3 Instead of putting the Euclidean metric on the space $\mathbb{R}^2$, we could have put, say, the **taxicab metric**

$$\Delta s = |\Delta x| + |\Delta y| \tag{11.1}$$

[17] More precisely, the two dimensional space we have been tacitly using, namely $\mathbb{R}^2$, is a *topological* space.

so named because this is how taxicabs in New York city measure the distance between points (number of city blocks). In this geometry, the shortest distance between two points is not a straight line in the Euclidean sense (instead it is a collection of zigs and zags through the city). But the resulting geometry is perfectly consistent.

11.4 Another metric one can put on the two dimensional plane is the **Lorentzian metric**

$$(\Delta s)^2 = (\Delta x)^2 - (\Delta y)^2 \tag{11.2}$$

Lorentzian geometry is quite different from Euclidean geometry. For example, in Lorentzian geometry the distance between two points can be imaginary!

11.5 In 1907 Hermann Minkowski observed that the natural geometry to use on spacetime (the collection of all possible events in the universe) is not Euclidean geometry, but rather Lorentzian geometry. [18] That is, Minkowski noted that, because the interval

$$\Delta\sigma := \sqrt{(\Delta x)^2 - (\Delta ct)^2} \tag{5.4}$$

is the same for all inertial observers, the natural "distance" between points on spacetime is the interval. Spacetime, equipped with the Lorentzian metric (5.4), is now called **Minkowski spacetime** or (inaccurately) **Minkowski space** for short. In this context, we call (5.4) the **Minkowski metric**, as it gives the "distance" between events in Minkowski spacetime.

The Erlangen Program

11.6 In the preceding sections we defined the idea of a geometry using a metric, or distance function, between points. But we could take a different approach, following the lead of the famous German mathematician Felix Klein. When Klein succeeded von Staudt as professor at the Philosophic Faculty of Erlangen University in 1872, he gave an inaugural lecture entitled "Comparative Consideration of Recent Geometric Researches". In the course of the lecture he set forth what later came to be known as his **Erlangen program**. The essence of Klein's program was the observation that one way to characterize a geometry is by means of the **transformations** that leave it invariant. For example, Euclidean geometry is characterized by the fact that distances between points and angles between

[18] This idea had been anticipated a year earlier by the famous French mathematician and philosopher, Henri Poincaré.

vectors are invariant under rotations and translations. Indeed, one can show that the Euclidean metric (5.1) is the only metric (up to a scalar multiple) that remains invariant under rotations and translations. So we could turn things around and **define** Euclidean geometry as that geometry which remains invariant under rotations and translations.

11.7 According to this philosophy, we ought to be able to define the Lorentzian geometry on Minkowski spacetime by means of a set of transformations that leave the Minkowski metric invariant. This can be done: the analogue of rotations in Euclidean space are precisely the Lorentz boosts in Minkowski spacetime! The entire set of symmetry transformations in four dimensional Minkowski spacetime includes translations, ordinary rotations, and Lorentz boosts. The rotations and boosts together are referred to collectively as **Lorentz transformations**. The entire collection is referred to as the **Poincaré group**. In the next section, we develop the geometric interpretation of Lorentz boosts.

Rotations Revisited

11.8 Consider two coordinate systems, one unprimed and one primed, that are rotated relative to one another:

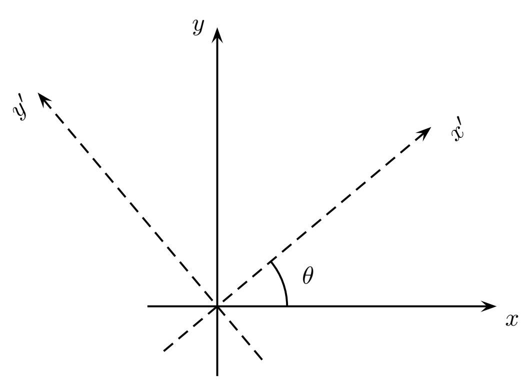


Figure 18. Two coordinate systems rotated relative to one another.

The coordinate transformation that takes us from the unprimed to the primed coordinates is given in matrix notation as follows: [19]

$$\boxed{\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}} \tag{11.3}$$

[19] As a convenience for the reader, the essentials of matrix algebra are given in Appendix B.

The equation (11.3) is shorthand for the following system of equations: [20]

$$
\begin{aligned}
x' &= x\cos\theta + y\sin\theta \\
y' &= -x\sin\theta + y\cos\theta
\end{aligned}
\tag{11.4}
$$

11.9 Example. The point represented by $(1,0)$ in the unprimed coordinates is represented in the primed coordinates by

$$
\begin{aligned}
\begin{pmatrix} x' \\ y' \end{pmatrix} &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta \\ -\sin\theta \end{pmatrix}
\end{aligned}
$$

which is correct:

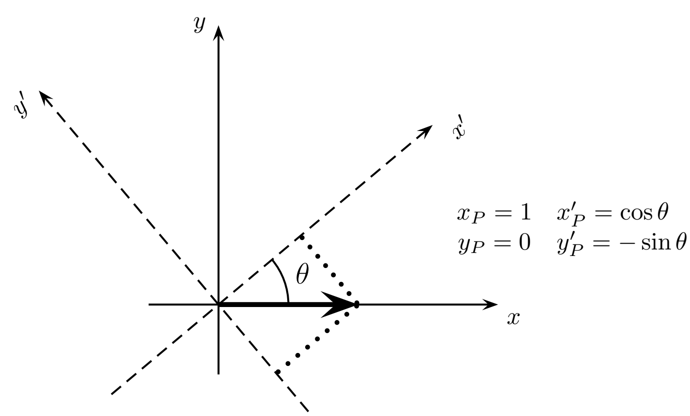


Figure 19. An example of a change of coordinates brought about by a rotation.

11.10 It is worth observing at this point that the inverse transformation, in which we go from the primed to the unprimed coordinates, is given by

$$
\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}
\tag{11.5}
$$

Written out explicitly, this becomes

$$
\begin{aligned}
x &= x'\cos\theta - y'\sin\theta \\
y &= x'\sin\theta + y'\cos\theta
\end{aligned}
\tag{11.6}
$$

You can verify that this is the correct inverse either by substituting (11.6) into (11.4) or else by using (11.5), (11.3), and matrix multiplication.

[20] If you are uncomfortable with the idea of matrices, you may always substitute (11.4) whenever you see an equation of the form (11.3).

11.11 Now, pick a point P in space represented by the coordinates (x, y) in the unprimed frame and (x', y') in the primed frame. These two pairs of coordinates are related by the transformations (11.3) and (11.5) (or (11.4) and (11.6)). We expect the length of the vector from the origin to P to be the same in both coordinate systems because a rotation cannot affect this length; the length is, if you will, **physical**, independent of the coordinates used to describe it. This is easy to verify:

$$\begin{aligned}(s')^2 = (x')^2 + (y')^2 &= (x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2 \\ &= x^2\cos^2\theta + 2xy\cos\theta\sin\theta + y^2\sin^2\theta \\ &\qquad + x^2\sin^2\theta - 2xy\sin\theta\cos\theta + y^2\cos^2\theta \\ &= (x^2+y^2)(\cos^2\theta + \sin^2\theta) \\ &= x^2 + y^2 = s^2 \end{aligned} \tag{11.7}$$

It is clear that the key fact that made the computation work is the trigonometric identity

$$\boxed{\cos^2\theta + \sin^2\theta = 1} \tag{11.8}$$

11.12 Because the transformation (11.3) is linear it follows that

$$\boxed{\begin{pmatrix}\Delta x' \\ \Delta y'\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}} \tag{11.9}$$

where $\Delta x = x_2 - x_1$ is the difference between the x coordinates of two different points, *etc.* .

It then follows, by a calculation entirely analogous to (11.7), that the distance Δs between any two points (x_1, y_1) and (x_2, y_2) is invariant under a rotation.

Lorentz Boosts

11.13 We now turn to the analogue of all this in Minkowski spacetime. We know that the transformation taking us from one coordinate system to another is given by the Lorentz boost (6.4) and its inverse (6.6). Written in matrix notation (with one event at the origin), we have

$$\boxed{\begin{pmatrix}ct' \\ x'\end{pmatrix} = \begin{pmatrix}\gamma_R & -\beta_R\gamma_R \\ -\beta_R\gamma_R & \gamma_R\end{pmatrix}\begin{pmatrix}ct \\ x\end{pmatrix}} \tag{11.10}$$

and

$$\boxed{\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} \gamma_R & \beta_R\gamma_R \\ \beta_R\gamma_R & \gamma_R \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix}} \tag{11.11}$$

11.14 We wish to write these transformations in a form that makes their geometrical significance manifest. The trick is to use **hyperbolic functions**: [21]

$$\boxed{\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\alpha_R & -\sinh\alpha_R \\ -\sinh\alpha_R & \cosh\alpha_R \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}} \tag{11.12}$$

Clearly, for this to represent the Lorentz transformation (11.10), we must have

$$\boxed{\cosh\alpha_R = \gamma_R} \tag{11.13}$$

and

$$\boxed{\sinh\alpha_R = \beta_R\gamma_R} \tag{11.14}$$

The parameter α_R is called the **velocity parameter**, because it is simply related to the velocity of the rocket:

$$\boxed{\tanh\alpha_R = \beta_R} \tag{11.15}$$

11.15 Just as (11.3) represents a **rotation** in **space**, (11.12) represents a **hyperbolic rotation** in **Minkowski spacetime**. And just as rotations preserve the Euclidean distance between points in space, so the hyperbolic rotation (Lorentz boost) (11.12) preserves the interval (Minkowski "distance") between events in spacetime. In the case of rotations, the key formula was (11.8). In the present case, the key formula is the hyperbolic analogue, namely

$$\boxed{\cosh^2\alpha_R - \sinh^2\alpha_R = 1} \tag{11.16}$$

as we can see from the following computation:

$$\begin{aligned}
(\sigma')^2 = (x')^2 - (ct')^2 &= (-ct\sinh\alpha_R + x\cosh\alpha_R)^2 - (ct\cosh\alpha_R - x\sinh\alpha_R)^2 \\
&= (ct)^2\sinh^2\alpha_R - 2ctx\sinh\alpha_R\cosh\alpha_R + x^2\cosh^2\alpha_R \\
&\qquad - (ct)^2\cosh^2\alpha_R + 2ctx\cosh\alpha_R\sinh\alpha_R - x^2\sinh^2\alpha_R \\
&= (x^2 - (ct)^2)(\cosh^2\alpha_R - \sinh^2\alpha_R) \\
&= x^2 - (ct)^2 = (\sigma)^2
\end{aligned}$$

[21] For a brief review of hyperbolic functions, consult Appendix C.

11.16 What does the transformation (11.12) do geometrically? The answer, of course, is that it is a hyperbolic rotation. But we cannot represent this rotation faithfully in Euclidean space, because Euclidean and Lorentzian geometry are different. We can, however, do the next best thing, which is to see what (11.12) does to spacetime diagrams. The presence of the two minus signs in the off-diagonal terms of the transformation matrix (unlike the rotation case) mean that the primed axes are related to the unprimed axes as follows:

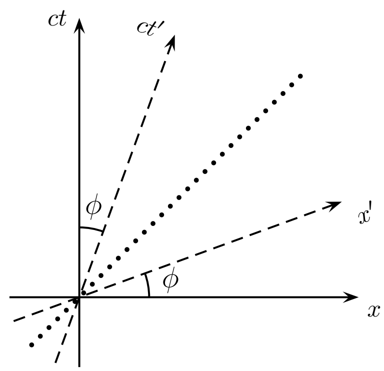


Figure 20. A geometrical representation of the coordinate transformation brought about by a Lorentz boost.

Prima facie, this behavior looks rather peculiar. For one thing, it does not look very symmetric. For example, an ordinary rotation could be viewed either as rotating the primed coordinate system counterclockwise through an angle θ or the unprimed coordinate system clockwise through the same angle. Either way, we get the same effect. Put another way, all we can measure are **relative** rotations, which is what we expect on physical grounds. Yet the Lorentz transformation looks decidedly asymmetric.

11.17 But this is simply an artifact of our limited representation of it. Indeed, were we to represent the angle between the t' and x' axes as a right angle, the t and x axes would be pictured as being rotated in the same way that the t' and x' axes are in Figure 20. The reason we cannot see this directly is that we are attempting to represent a **hyperbolic** rotation in **Euclidean** space. The result is necessarily something of a compromise. This is not to say that Figure 20 is not useful. It does tell us many things about the two observers. First, the angle ϕ between the x and x' (or ct and ct') axes is just $\tan^{-1}\beta_R$, where β_R is the relative speed of the two observers.

11.18 Exercise. Prove this.

This means that the angle ϕ of rotation is related to the velocity parameter α_R. In fact, we must have

$$\tan\phi = \beta_R = \tanh\alpha_R \tag{11.17}$$

Second, notice that the line of symmetry $ct' = x'$ of the rocket observer coincides precisely with the line of symmetry $ct = x$ of the laboratory observer.

11.19 EXERCISE. Explain the physical significance of this last fact.

11.20 We may use this formalism to develop a geometric interpretation of the relativistic velocity addition formula (9.5). One can view that equation as giving the result of performing a Lorentz boost **twice**—once to transform coordinates to the frame of the rocket, and once more to transform to the frame of the baseball. And, as we discovered, the velocities of the two frames do not simply add (*cf.* (9.5)). But there is a very natural parameter that **does** simply add: the velocity parameter α_R! To see this, recall from (11.15) that the velocity parameter is defined *via* the relation $\tanh\alpha_R = \beta_R$. So, defining α' and α analogously, we find, from (9.5):

$$\tanh\alpha = \frac{\tanh\alpha_R + \tanh\alpha'}{1 + \tanh\alpha_R \tanh\alpha'} \tag{11.18}$$

Comparing this to the addition formula for hyperbolic tangent in Appendix C (Equation ***C*.13**) we see that

$$\boxed{\alpha = \alpha_R + \alpha'} \tag{11.19}$$

11.21 This relation brings out quite clearly the **geometrical significance** of successive Lorentz boosts: Just as successive rotations are obtained by adding the angles θ of rotation, successive Lorentz boosts are obtained by adding the velocity parameters α of the boost!

12. THE POSITION 2-VECTOR

12.1 Although it may be more elegant and geometrically suggestive, the hyperbolic representation of Lorentz boosts is not the best formalism to use for everyday computations. In this Chapter we continue our development of the geometry of Minkowski spacetime using the β and γ parameters introduced previously. We begin by formalizing some of the ideas we have encountered informally in the previous sections.

12.2 In Chapter **5** and again in Chapter **11** we located each event in spacetime by means of its coordinates ct and x. Putting them together we may define the **position 2-vector** $\widetilde{\boldsymbol{x}}$ of the event as follows:

$$\boxed{\widetilde{\boldsymbol{x}} := (x^0, x^1) = (ct, x) \equiv \begin{pmatrix} ct \\ x \end{pmatrix}} \tag{12.1}$$

This is just like an ordinary vector in space, except it is in spacetime. Observe that both components have the same units, namely distance. In spite of this, the zeroth component x^0 is called the **time component** and the first component x^1 is called the **space component** of the 2-vector $\widetilde{\boldsymbol{x}}$. [22]

12.3 As we learned in Chapter **11**, different observers will, in general, assign different position 2-vectors to the same event. But, provided the origins of both reference frames coincide at $t = 0$, their respective 2-vectors will be related by a Lorentz boost (11.10). If we denote the Lorentz boost matrix by

$$L := \begin{pmatrix} \gamma_R & -\beta_R\gamma_R \\ -\beta_R\gamma_R & \gamma_R \end{pmatrix} \tag{12.2}$$

then Equation (11.10) may be written

$$\boxed{\widetilde{\boldsymbol{x}}' = L\widetilde{\boldsymbol{x}}} \tag{12.3}$$

12.4 Next we define the **length** of the position 2-vector $\widetilde{\boldsymbol{x}}$ to be simply the **interval** between the origin (event $(0, 0)$) and the event whose spacetime coordinates are (ct, x):

$$|\widetilde{\boldsymbol{x}}| = \sqrt{x^2 - (ct)^2} = \sigma \tag{12.4}$$

We know, of course, that this is a **relativistic invariant**; that is, it is the same for all observers: $|\widetilde{\boldsymbol{x}}| = |\widetilde{\boldsymbol{x}}'|$.

[22] A word of caution is in order here. The expression x^0, for example, means the time component of the 2-vector $\widetilde{\boldsymbol{x}}$. It does **not** mean the number x raised to the power 0. You might ask why we do not avoid this problem by writing x_0 instead of x^0. The problem is that these two quantities mean different things! In special relativity (and, *a fortiori*, in general relativity) the placement of the indices matters. In Minkowski spacetime, for example, $x_0 = -x^0$.

12.5 We have been restricting ourselves to $1+1$ dimensions. In $3+1$ dimensions (the real world), we have the **position 4-vector**

$$\widetilde{\boldsymbol{x}} = (x^0, x^1, x^2, x^3) = (ct, x, y, z) = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{12.5}$$

and, for a Lorentz boost in the x direction (rocket ship traveling in the $+x$ direction), we have

$$L := \begin{pmatrix} \gamma_R & -\beta_R\gamma_R & 0 & 0 \\ -\beta_R\gamma_R & \gamma_R & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{12.6}$$

For our purposes, however, we may continue to restrict ourselves to the artificial world of one spatial and one temporal dimension.

13. 2-Vectors

13.1 You may be wondering why I did not simply call $\widetilde{\boldsymbol{x}}$ a 'vector' instead of a '2-vector'. The reason has to do with the meaning carried by the word 'vector'. To a mathematician, a vector is merely some object that lives in a vector space (*cf.* Appendix A). To a physicist, however, **a vector is an object that transforms like the position vector under rotations**. What does this mean?

13.2 As we saw in Chapter **11**, under a rotation the position vector of an event transforms according to equation (11.3). A quantity $\boldsymbol{q} = (q^x, q^y)$ is called a **vector** provided it transforms the same way under a rotation:

$$\begin{pmatrix} q'^x \\ q'^y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} q^x \\ q^y \end{pmatrix} \tag{13.1}$$

For example, the velocity of a particle is a vector because, under a rotation, the components of the velocity vector transform according to (13.1). On the other hand, suppose I were to assemble the position x and momentum p_x of a particle into a two component object (x, p_x). Then, even though we could add objects such as these using our usual rules of componentwise addition, we could not call (x, p_x) a vector, because it fails to satisfy the transformation rule (13.1). [23] [24]

13.3 By definition, **a 2-vector is an object that transforms like the position 2-vector under Lorentz transformations**. Thus, a two-component quantity $\widetilde{\boldsymbol{\xi}} = (\xi^0, \xi^1)$ is a two-vector provided it transforms according to the rule

$$\boxed{\widetilde{\boldsymbol{\xi}}' = L\widetilde{\boldsymbol{\xi}}} \tag{13.2}$$

where L is the Lorentz boost matrix (12.2). As before, not every two-component object in spacetime is a 2-vector. But the condition of being a 2-vector is an important one, for if you know that something is a 2-vector then you immediately know its behavior under Lorentz transformations.

13.4 We define the **length** of any 2-vector to be given by the expression

$$\boxed{|\widetilde{\boldsymbol{\xi}}| = \sqrt{(\xi^1)^2 - (\xi^0)^2}} \tag{13.3}$$

(this is just the Minkowski metric). It follows that, if you know that a physical quantity is a 2-vector, then you immediately know that its length is a Lorentz scalar. That is, **the length of a 2-vector is invariant under Lorentz boosts.** The computation is exactly the same as the one performed in Section **11.15**.

[23] We say instead that x and p_x transform as the x-components of two different vectors, namely the position vector (x, y) and the momentum vector (p_x, p_y), respectively.

[24] Actually, there is one more subtlety. We also make a distinction between **polar vectors** (often just called **vectors**!) and **axial vectors** or **pseudovectors**. A polar vector is one which transforms under an **inversion** (**parity transformation**) like the position vector. Under an inversion the position vector $\boldsymbol{r} = (x, y, z)$ goes to the vector $(-x, -y, -z) = -(x, y, z) = -\boldsymbol{r}$. Hence, for example, the velocity vector is a polar vector, because it is defined as $\boldsymbol{v} = d\boldsymbol{r}/dt$, and nothing happens to time if space is inverted. By contrast, the angular momentum vector is a pseudovector, because, although it transforms as does the position vector under rotations, under inversions it remains invariant. This follows because $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} = \boldsymbol{r} \times m\boldsymbol{v}$ and both $\boldsymbol{r}$ and $\boldsymbol{v}$ change sign under inversion.

Incidentally, the quantity $\boldsymbol{L}\cdot\boldsymbol{L}$ (the dot product of $\boldsymbol{L}$ with itself) is a **scalar** because it is invariant under rotations (and inversions). But the quantity $\boldsymbol{r}\cdot\boldsymbol{L}$ is a **pseudoscalar** because, while it is invariant under rotations, it changes sign under inversions. The importance of all this lies in the fact that the transformation properties of physical quantities oftentimes provide additional data that help us deduce the correct laws of physics. For example, if it is known that parity is conserved in a certain process, then the laws governing that process will have to be parity invariant. This, in turn, puts restrictions on the manner in which various quantities (*e.g.,* scalar, pseudoscalar, vector, pseudovector) appear in the equations.

13.5 Exercise. Prove this.

This result is, of course, not a surprise. For, as we discussed in Section **11.7**, the Lorentz transformations consist precisely of those linear transformations that leave the length of a 2-vector invariant.

13.6 As noted in Section **12.5**, the real world has 3 spatial dimensions. So to be correct we should really speak of **4-vectors**, which are defined analogously as 4-component objects $\widetilde{\boldsymbol{\xi}} = (\xi^0, \xi^1, \xi^2, \xi^3)$ that transform like the position 4-vector under Lorentz transformations (*cf.* Equation (12.6)). Similarly, one defines the length of a 4-vector to be

$$|\widetilde{\boldsymbol{\xi}}| = \sqrt{-(\xi^0)^2 + (\xi^1)^2 + (\xi^2)^2 + (\xi^3)^2}$$

It is invariant under all Lorentz transformations (which, you will recall, includes both rotations and boosts).

14. The Velocity 2-Vector

14.1 We know that the ordinary velocity vector is obtained by differentiating the ordinary position vector with respect to time. We wish to do something similar in Minkowski space, but we will find that we must modify this idea slightly in order to obtain a definition that makes sense from the point of view of relativity. Suppose we have two events in spacetime, located by their respective position 2-vectors $\widetilde{\boldsymbol{x}}_1$ and $\widetilde{\boldsymbol{x}}_2$. Then the **displacement 2-vector** between them is exactly what you would expect it to be, namely

$$\Delta\widetilde{\boldsymbol{x}} := \widetilde{\boldsymbol{x}}_2 - \widetilde{\boldsymbol{x}}_1 = (c(t_2 - t_1), x_2 - x_1) = (c\Delta t, \Delta x) = \begin{pmatrix} c\Delta t \\ \Delta x \end{pmatrix} \tag{14.1}$$

This is illustrated in Figure 21 below:

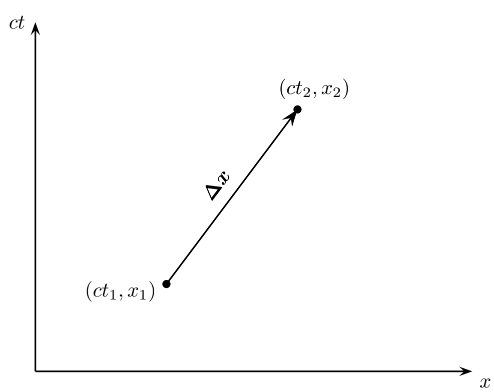


Figure 21. The displacement 2-vector.

Note that the displacement 2-vector is indeed a 2-vector, by the linearity of (12.3). Two different observers moving with relative velocity β_R compute different displacement 2-vectors between the same two events; they are related by

$$\Delta\widetilde{\boldsymbol{x}}' = L\Delta\widetilde{\boldsymbol{x}} \tag{14.2}$$

where L is given by Equation (12.2). The length of the displacement 2-vector is precisely the interval between the two events

$$|\Delta\widetilde{\boldsymbol{x}}| = \sqrt{(\Delta x)^2 - (c\Delta t)^2} = \Delta\sigma \tag{14.3}$$

14.2 Now, consider a particle, which may be moving or at rest. Its history (namely, the temporal succession of points in space that it visits) is represented on a spacetime diagram by its **worldline** (see Figure 22).

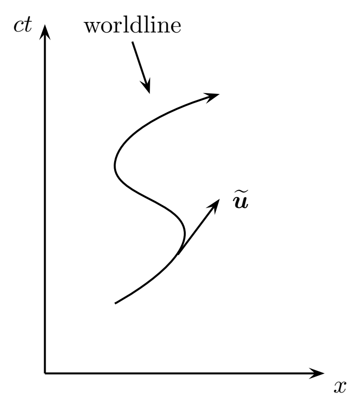


Figure 22. The velocity 2-vector is tangent to the particle's worldline.

We may mark off the particle's passage along its worldline by means of a clock carried by the particle. As we learned, the clock carried by the particle measures **proper time**. So it is natural to define the **velocity 2-vector** of the particle as the tangent vector to the curve, *i.e.*, the derivative of the position 2-vector with respect to proper time:

$$\boxed{\widetilde{\boldsymbol{u}} := \lim_{\Delta\tau\to 0} \frac{\Delta\widetilde{\boldsymbol{x}}}{\Delta\tau} = \frac{d\widetilde{\boldsymbol{x}}}{d\tau}} \tag{14.4}$$

It follows from the definition (14.1) that

$$\widetilde{\boldsymbol{u}} = \lim_{\Delta\tau\to 0}(c\frac{\Delta t}{\Delta\tau}, \frac{\Delta x}{\Delta\tau}) = (c\frac{dt}{d\tau}, \frac{dx}{d\tau}) = (u^0, u^1) \tag{14.5}$$

14.3 Of course, we have no right to call this a 2-vector until we demonstrate that it transforms under Lorentz boosts in the correct way. But this is clear from the definition. As we argued above, $\Delta\widetilde{\boldsymbol{x}}$ is a 2-vector, so by continuity $d\widetilde{\boldsymbol{x}}$ is also a 2-vector. Now $\Delta\tau$ is a Lorentz scalar (invariant), so $d\tau$ is also a Lorentz scalar. It follows that $\widetilde{\boldsymbol{u}} = d\widetilde{\boldsymbol{x}}/d\tau$ is indeed a 2-vector, transforming according to the rule

$$\boxed{\widetilde{\boldsymbol{u}}' = L\widetilde{\boldsymbol{u}}} \tag{14.6}$$

Written out in full, this reads

$$\begin{pmatrix} u'^0 \\ u'^1 \end{pmatrix} = \begin{pmatrix} \gamma_R & -\beta_R\gamma_R \\ -\beta_R\gamma_R & \gamma_R \end{pmatrix} \begin{pmatrix} u^0 \\ u^1 \end{pmatrix} \tag{14.7}$$

14.4 As $\widetilde{\boldsymbol{u}}$ is a 2-vector, we know immediately that its length must be a relativistic invariant. But we may see this directly as follows:

$$|\widetilde{\boldsymbol{u}}| = \left|\frac{d\widetilde{\boldsymbol{x}}}{d\tau}\right| = \frac{|d\widetilde{\boldsymbol{x}}|}{|d\tau|} = \frac{ic\,d\tau}{d\tau} = ic \tag{14.8}$$

where we used the infinitessimal versions of (5.6) and (14.3). As all observers agree on the values of i and c, it follows from the principles of relativity that all observers will agree on the length of the velocity 2-vector of any particle:

$$|\widetilde{\boldsymbol{u}}'| = |\widetilde{\boldsymbol{u}}| = ic \tag{14.9}$$

The physical significance of this result will become clear in Chapter **15**.

14.5 It will be useful to have a more explicit form for the velocity 2-vector of a particle. To that end, using the infinitessimal form of (5.5), we observe that

$$\frac{dt}{d\tau} = \frac{dt}{\sqrt{(dt)^2 - (dx/c)^2}} = \frac{1}{\sqrt{1-(dx/c\,dt)^2}} \tag{14.10}$$

Next, define v to be the ordinary velocity of the particle as measured by the laboratory observer: [25]

$$\boxed{v := \frac{dx}{dt}} \tag{14.11}$$

[25] Recall that the speed of the rocket relative to the lab is denoted v_R.

We then make the definitions

$$\boxed{\beta := \frac{v}{c} \qquad \text{and} \qquad \gamma := \frac{1}{\sqrt{1-\beta^2}}} \tag{14.12}$$

with the reminder not to confuse this β and γ with β_R and γ_R. The β without any subscript refers to the velocity of a particle **in a fixed reference frame**. The symbol β_R refers to the velocity of one reference frame relative to another. Ordinarily these two quantities have nothing to do with one another, so try to keep them straight.

14.6 Combining (14.10), (14.11), and (14.12) we have

$$\frac{dt}{d\tau} = \gamma \tag{14.13}$$

as well as

$$\frac{dx}{d\tau} = \frac{dx}{dt}\frac{dt}{d\tau} = c\beta\gamma \tag{14.14}$$

whence Equation (14.5) becomes

$$\widetilde{\boldsymbol{u}} = (c\gamma, c\beta\gamma) = c\gamma(1,\beta) = \gamma(c,v) = (u^0, u^1) \tag{14.15}$$

14.7 In 3 spatial dimensions the ordinary velocity vector $\boldsymbol{v}$ is defined to be

$$\boldsymbol{v} := \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) \tag{14.16}$$

so that the three dimensional analogue of (14.12) is

$$\boldsymbol{\beta} = \frac{\boldsymbol{v}}{c} \qquad \text{and} \qquad \gamma = \frac{1}{\sqrt{1-\boldsymbol{\beta}\cdot\boldsymbol{\beta}}} \tag{14.17}$$

The **velocity 4-vector** is given by

$$\widetilde{\boldsymbol{u}} = \frac{d\widetilde{\boldsymbol{x}}}{dt} = c(1,\boldsymbol{\beta}) = c\gamma(1,\beta^x,\beta^y,\beta^z) = (u^0,u^1,u^2,u^3) \tag{14.18}$$

15. The Energy-Momentum 2-Vector

15.1 By analogy with the definition of momentum in non-relativistic mechanics, we define the **energy-momentum 2-vector** of a particle as follows:

$$\boxed{\widetilde{\boldsymbol{p}} = m\widetilde{\boldsymbol{u}}} \tag{15.1}$$

where m is the mass of the particle (about which we shall have more to say later) and $\widetilde{\boldsymbol{u}}$ is the particle's velocity 2-vector. Once again, we really have no business calling this a '2-vector' unless we can show that it is indeed a 2-vector. Fortunately, we shall see that it does indeed satisfy the requisite properties.

15.2 Before we verify this, though, let us justify the terminology of 'energy-momentum'. Using (14.15) and (15.1) we may write

$$\widetilde{\boldsymbol{p}} = mc\gamma(1, \beta) \tag{15.2}$$

Consider the space component of $\widetilde{\boldsymbol{p}}$ when the particle velocity is small. In that case we may Taylor expand γ as in (4.9) to get (to lowest order in the particle velocity)

$$mc\beta\gamma = mv\gamma \approx mv(1 + \beta^2/2) \approx mv = p_{\text{non-relativistic}} \tag{15.3}$$

We see that the space component of the energy-momentum 2-vector reduces to the ordinary non-relativistic momentum of the particle when its velocity is small. Because of this, we define the **relativistic momentum** (or simply **momentum**) of the particle to be

$$\boxed{p = mc\beta\gamma = mv\gamma} \tag{15.4}$$

15.3 This justifies half of the name 'energy-momentum'. To justify the other half, consider the time component of $\widetilde{\boldsymbol{p}}$, again in the limit of low particle velocity:

$$\begin{aligned} mc\gamma &= mc(1 + \beta^2/2 + ...) \\ &= \frac{1}{c}[mc^2 + \frac{1}{2}mc^2\beta^2 + ...] \\ &= \frac{1}{c}[mc^2 + \frac{1}{2}mv^2 + ...] \end{aligned} \tag{15.5}$$

Examination of the second term in the brackets reveals that it is nothing more than the ordinary non-relativistic kinetic energy of the particle. Furthermore, as all the higher order terms involve higher powers of β^2, they vanish when the particle is at rest ($\beta = 0$). When this happens, the first term mc^2 remains. For this reason, the first term is called the **rest energy** of the particle, because it is an energy and it is associated to the particle when the particle is at rest.

15.4 It therefore seems natural to call the entire quantity in the brackets in (15.5) E, the **total relativistic energy of the particle**:

$$\boxed{E := mc^2\gamma} \tag{15.6}$$

Note that when the particle is at rest, $\gamma = 1$ and this expression reduces to the famous expression

$$E_{\text{rest}} = mc^2 \tag{15.7}$$

expressing the equivalence of energy and mass.

15.5 It is worth observing that Equation (15.6) provides yet another reason why no one can travel at the speed of light. As the particle approaches the speed of light, γ approaches infinity, so E approaches infinity. In other words, it takes an infinite amount of energy to accelerate a particle to the speed of light, which is impossible (the universe does not have this much energy).

15.6 One of the consequences of (15.7) is that mass is simply a kind of energy: if you heat an object, or hit it, or compress it, or in any way add energy to it, you increase its mass accordingly.

15.7 We define the **relativistic kinetic energy** T of a particle as the energy the particle has by virtue of its motion, namely

$$\boxed{T := E - mc^2 = mc^2(\gamma - 1)} \tag{15.8}$$

15.8 It follows from (15.2), (15.4) and (15.6) that

$$\boxed{\widetilde{\boldsymbol{p}} = (E/c, p) = (p^0, p^1)} \tag{15.9}$$

where E and p are the total relativistic energy and momentum, respectively, of the particle. It is for this reason that $\widetilde{\boldsymbol{p}}$ is called the **energy-momentum** 2-vector of the particle.

15.9 You may object that we are simply **defining** the momentum and energy of the particle. At this point the only motivation for our definitions is that they are plausible generalizations of the familiar non-relativistic quantities that reduce to them in the appropriate limit, and that they fit together nicely into a 2-vector (which we have yet to prove). If you go back to Newtonian mechanics you will see that we simply defined momentum and energy there as well. The utility of their definition in that case was that they were **conserved** in all mechanical processes. Well, the same is true here, but a reasonably rigorous demonstration of this starting from the conservation laws and the postulates of relativity requires a bit of computation. [26]

15.10 The approach we chose, namely to simply define $\widetilde{\boldsymbol{p}} = m\widetilde{\boldsymbol{u}}$, has two advantages over the more pedestrian approach. First, it is considerably more succinct, and second, it conforms to the more modern philosophy of postulating the relativistic generalization of non-relativistic quantities based on the desire to make the **Lorentz covariance** of the laws of physics manifest. Lorentz covariance means that, under a Lorentz transformation the various dynamical observables change in such a way as to leave the **law** relating them invariant (Postulate 1). When the laws of physics are expressed in terms of 4-vectors, or, more generally, **tensors**, their Lorentz covariance properties are immediately apparent.

15.11 For example, Maxwell's two equations with sources may be written $\widetilde{\boldsymbol{G}} = (4\pi/c)\widetilde{\boldsymbol{J}}$, where $\widetilde{\boldsymbol{G}}$ is some 4-vector that depends on the electric and magnetic fields, and $\widetilde{\boldsymbol{J}}$ is the 4-vector built from the charge and current densities. The point is that, written like this in terms of 4-vectors, it is immediately obvious that these two equations are Lorentz covariant. This is because, under a Lorentz transformation the equations become $\widetilde{\boldsymbol{G}}' = (4\pi/c)\widetilde{\boldsymbol{J}}'$, where $\widetilde{\boldsymbol{G}}' = L\widetilde{\boldsymbol{G}}$ and $\widetilde{\boldsymbol{J}}' = L\widetilde{\boldsymbol{J}}$. That is, the **law** relating fields to sources remains the same in all inertial reference frames. As an aside, observe that we could write the transformed Maxwell equations as $L(\widetilde{\boldsymbol{G}} - (4\pi/c)\widetilde{\boldsymbol{J}}) = 0$. This equation holds if and only if $\widetilde{\boldsymbol{G}} = (4\pi/c)\widetilde{\boldsymbol{J}}$ (because L is an arbitrary Lorentz transformation), which is another way of expressing the fact that, if the law holds in one reference frame, it holds in all of them.

15.12 Finally, we must justify the terminology '2-vector' for $\widetilde{\boldsymbol{p}}$. To do so, we must first

[26] See, for example, J.D. Jackson, *Classical Electrodynamics* (John Wiley, New York, 1975) Chapter 11.

discuss the 'm' that appears in (15.1). We call this the **rest mass** of the particle, and **require** that **the rest mass of a particle be a relativistic invariant**. That is, we demand that all observers measure the same value for m. This is certainly in accord with our old Newtonian intuitions (but see Section **15.16** below).

15.13 Given that m is a relativistic invariant, it is now immediately obvious that $\widetilde{\boldsymbol{p}}(= m\widetilde{\boldsymbol{u}})$ is a 2-vector: m does not change under a Lorentz transformation, while $\widetilde{\boldsymbol{u}}$ transforms like a 2-vector. Hence $\widetilde{\boldsymbol{p}}$ is a 2-vector, transforming according to the rule:

$$\boxed{\widetilde{\boldsymbol{p}}' = L\widetilde{\boldsymbol{p}}} \tag{15.10}$$

Written out in full, this reads:

$$\begin{pmatrix} E'/c \\ p' \end{pmatrix} = \begin{pmatrix} \gamma_R & -\beta_R\gamma_R \\ -\beta_R\gamma_R & \gamma_R \end{pmatrix} \begin{pmatrix} E/c \\ p \end{pmatrix} \tag{15.11}$$

It follows from (15.1) that

$$|\widetilde{\boldsymbol{p}}| = |m\widetilde{\boldsymbol{u}}| = m|\widetilde{\boldsymbol{u}}| = imc \tag{15.12}$$

which makes it immediately obvious that

$$|\widetilde{\boldsymbol{p}}'| = |\widetilde{\boldsymbol{p}}| \tag{15.13}$$

i.e., that the length of the energy-momentum 2-vector is a relativistic invariant.

15.14 We can deduce an important consequence of (15.13) if we use the definition (13.3) of the length of a 2-vector (applied to $\widetilde{\boldsymbol{p}}$):

$$imc = |\widetilde{\boldsymbol{p}}| = \sqrt{(p^1)^2 - (p^0)^2} = \sqrt{p^2 - (E/c)^2} \tag{15.14}$$

Squaring both sides and rearranging we find

$$\boxed{E^2 = (pc)^2 + (mc^2)^2} \tag{15.15}$$

which is referred to as the **invariant mass equation**. It expresses the fact that the rest mass of a particle is a relativistic invariant, and shows that the relativistic energy and relativistic momentum of a particle are not independent.

15.15 For completeness, let us extend these ideas to 3+1 dimensions. Momentum now becomes a vector

$$\boldsymbol{p} := m\boldsymbol{v}\gamma = m\gamma(v^x, v^y, v^z) \tag{15.16}$$

so we have the **energy-momentum 4-vector**:

$$\widetilde{\boldsymbol{p}} := (E/c, \boldsymbol{p}) = (p^0, p^1, p^2, p^3) \tag{15.17}$$

It transforms according to the rule

$$\widetilde{\boldsymbol{p}}' = L\widetilde{\boldsymbol{p}} \tag{15.18}$$

where L is now a Lorentz transformation in 3+1 dimensions (*cf.* (12.6)). Equation (15.15) still holds, with $p^2 = \boldsymbol{p}\cdot\boldsymbol{p}$, as you should verify. (Be careful not to confuse this p^2 with the second component of $\widetilde{\boldsymbol{p}}$!)

The Unfortunate Idea of Relativistic Mass

15.16 Some people, merely because they do not like the expression (15.16) for the momentum of a particle, define a new quantity m^* which they call the **'relativistic mass'** of a particle. It is defined according to the formula

$$m^* = m\gamma \tag{15.19}$$

Observe that the 'relativistic mass' of a particle **is not a relativistic invariant**. Instead, the 'relativistic mass' of a particle grows without bound as the particle approaches light-speed. With this definition, the total relativistic energy and momentum of a particle may be written

$$E = m^* c^2 \qquad \text{and} \qquad \boldsymbol{p} = m^* \boldsymbol{v} \tag{15.20}$$

When written this way, the second equation looks more like the old familiar equation for momentum.

15.17 Well, to be fair, there is another reason why people introduce the idea of 'relativistic mass'. It is that the equation for relativistic momentum written in the form $\boldsymbol{p} = m^*\boldsymbol{v}$ may be viewed as expressing the (true) fact that the particle becomes progressively harder to deflect as its velocity increases. In ordinary non-relativistic mechanics, it takes just as much force to deflect a particle when it is moving as when it is at rest. This is definitely not true when the particle approaches light speed.

15.18 Even though the computation is rather ugly, I will demonstrate this claim. [27] In our reference frame, we define the concept of force using Newton's 2nd Law. So the force vector

[27] Adapted from J. Rohlf, *Modern Physics from α to Z^0* (John Wiley, New York, 1994).

we must apply to an object to change its momentum $\boldsymbol{p}$ in a time dt is simply $\boldsymbol{F} = d\boldsymbol{p}/dt$. If we substitute the correct (relativistic) expression for the particle's momentum into this equation we get

$$\boldsymbol{F} = \frac{d(m\boldsymbol{v}\gamma)}{dt} = m\frac{d(\boldsymbol{v}\gamma)}{dt} = m\gamma\frac{d\boldsymbol{v}}{dt} + m\boldsymbol{v}\frac{d\gamma}{dt}$$

But

$$\frac{d\gamma}{dt} = (1-\beta^2)^{-3/2}\boldsymbol{\beta}\cdot\frac{d\boldsymbol{\beta}}{dt} = \gamma^3\boldsymbol{\beta}\cdot\frac{d\boldsymbol{\beta}}{dt}$$

Observe that

$$\begin{aligned}\boldsymbol{F}\cdot\boldsymbol{v} &= m\gamma\boldsymbol{v}\cdot\frac{d\boldsymbol{v}}{dt} + mv^2\gamma^3\boldsymbol{\beta}\cdot\frac{d\boldsymbol{\beta}}{dt}\\ &= m\gamma\boldsymbol{v}\cdot\frac{d\boldsymbol{v}}{dt} + m\beta^2\gamma^3\boldsymbol{v}\cdot\frac{d\boldsymbol{v}}{dt}\\ &= m\gamma\boldsymbol{v}\cdot\frac{d\boldsymbol{v}}{dt}(1+\gamma^2\beta^2)\\ &= m\gamma^3\boldsymbol{v}\cdot\frac{d\boldsymbol{v}}{dt}\end{aligned}$$

so that

$$\frac{d\gamma}{dt} = \frac{\boldsymbol{F}\cdot\boldsymbol{v}}{mc^2}$$

in which case the expression for the force becomes

$$\boldsymbol{F} = m\gamma\frac{d\boldsymbol{v}}{dt} + \boldsymbol{\beta}(\boldsymbol{F}\cdot\boldsymbol{\beta})$$

and the acceleration of the particle is

$$\boldsymbol{a} := \frac{d\boldsymbol{v}}{dt} = \frac{\boldsymbol{F} - \boldsymbol{\beta}(\boldsymbol{F}\cdot\boldsymbol{\beta})}{m\gamma}$$

If the force $\boldsymbol{F}$ is applied in a direction orthogonal to the direction of motion of the particle, this equation reduces to the familiar form of Newton's 2nd Law

$$\boldsymbol{F} = m^*\boldsymbol{a}$$

This makes it clear that the **inertial mass** of the particle is velocity dependent. It is for this reason that some people wish to assert that the mass of the particle increases with increasing velocity.

15.19 We refuse to succumb to such sloppy thinking, as should you. Not only is the concept of relativistic mass distasteful, it is also dangerous. It is distasteful because it contravenes the essence of relativity, which is to search for those physical quantities that remain invariant under Lorentz transformations—*i.e.,* that remain the same for all moving observers. It is dangerous because it gives the mistaken impression that mass increases with velocity, **which is patently false**. Captain Kirk on the Starship Enterprise does **not** get heavier as he travels to Beta Eridani 3 (unless he eats too much Aldebaran stew). Nothing strange happens to matter as it moves faster and faster. In its own rest frame, a particle's mass does not change.

15.20 As an example of the sort of fallacy which one might commit if one holds the view that mass changes with speed, consider the following paradox. The Starship Enterprise moves close to the speed of light. At that speed, so the argument goes, its mass is so large (and its length so small) that the entire ship lies within its own Schwarzschild radius—in other words, it collapses to form a black hole! On the other hand, from the point of view of someone on the bridge of the Enterprise (which is, after all, a perfectly legitimate viewpoint), it is the **entire universe** that collapses to form a black hole. This is a true paradox, because one could easily tell which one of the two was the black hole. The paradox is resolved by the observation that the rest mass of a particle is a relativistic invariant, and it is the rest mass that determines whether an object will or will not form a black hole. This is fortunate for us, for it would be most inconvenient if, every time a particle were accelerated near light speed, the entire universe were to collapse!

15.21 Henceforth, we banish the phrase 'relativistic mass', and use the term 'mass' to mean 'rest mass'. It is the only mass that has any invariant (and therefore physical) meaning.

Zero Rest Mass

15.22 One consequence of our discussion is the following. As we observed previously, any particle that moves with the speed of light must have an infinite energy, as $E = mc^2\gamma$, and γ is infinite when $\beta = 1$. For this reason, we argued, no material object can travel at the speed of light. But what about **photons**? A photon is a particle of light, which, by definition, travels at the speed of light. Well, to avoid the absurd conclusion that the photon carries an infinite energy, we are forced to admit that **the photon has zero rest mass**! This may seem a little strange at first, for what would happen if we were to bring the photon to rest? Well, the obvious answer is that it would disappear! This is in fact true. You cannot stop a photon. It either travels at the speed of light, or it does not travel at all. For future use we note that, according to equation (15.15), the energy of a photon must be related to the momentum of the photon by the equation [28]

$$\boxed{E = pc} \tag{15.21}$$

16. Binding Energies

16.1 Before we can apply our newly acquired knowledge to the study of relativistic kinematics, we must discuss the appropriate units to use for subatomic phenomena.

Definition. One **unified atomic mass unit** (abbreviated 1 u) equals one twelfth of the mass of a neutral carbon-12 **atom**. [29]

The utility of this definition is that a typical nucleus has mass number A roughly equal to its mass in unified atomic mass units. For example, the mass number of the cesium isotope ^{137}Cs is 137. The nucleus of this element contains 55 protons and 82 neutrons for a total of 137 nucleons (its mass number). Its atomic mass is 136.907073 u, which rounds off numerically to 137 u. Put another way, the mass of a nucleon is approximately 1 u (although this ignores the electrons and the binding energy—see below).

As one **mole** of ^{12}C contains N_A (Avogadro's number) atoms and has a mass of 12 g, we have

$$1\,\mathrm{u} = \frac{1\,\mathrm{g}}{6.022 \times 10^{23}\ \text{particles/mole}} = 1.6605 \times 10^{-24}\,\mathrm{g} = 1.6605 \times 10^{-27}\,\mathrm{kg}$$

16.2 Recall that one electron volt (1 eV) is defined as the energy required to carry an electron through a potential difference of 1 volt. We therefore have

$$1\,\mathrm{eV} = (1.6 \times 10^{-19}\,\mathrm{C})(1\,\mathrm{V}) = 1.6 \times 10^{-19}\,\mathrm{J}$$

[28] I like to say that the energy of a photon is "politically correct". This is an obscure reference to a bizarre ideology that achieved short-lived prominence in late 20th century America, in which some extremists tried to ban various sorts of oral and written phrases that might offend anyone in some way. Thanks to the efforts of a few clear thinking citizens, this practice was soon abandoned.

[29] This definition supersedes the old definition of the **atomic mass unit** (abbreviated 1 amu), which was defined as one sixteenth of the mass of a neutral oxygen-16 atom.

This allows us to compute the energy equivalent of 1 u according to the Einstein formula (15.7)

$$E = mc^2 = \frac{(1.6605 \times 10^{-27}\,\mathrm{kg})(2.9979 \times 10^{8}\,\mathrm{m/s})^2}{1.6022 \times 10^{-13}\,\mathrm{J/MeV}} = 931.5\,\mathrm{MeV}$$

16.3 Next we have the following

Definition. The **binding energy** of a composite system is the energy required to break it up into its constituent parts.

In other words, to break something up into pieces, you must hit it with a hammer—you must add energy to the system. This seems obvious, but it has somewhat surprising implications, one of which is that the pieces of a system weigh more than the system itself! We can see this by considering the following

16.4 Example. The **deuteron** is a bound state of the proton and neutron. Measurements reveal that the mass of the neutron is $m_n = 1.008665$ u, while the mass of a hydrogen atom is 1.007825 u and the mass of deuterium (the atom of which the deuteron is the nucleus) is 2.014102 u. The latter two masses include the mass (and binding energy, which we ignore because it is so small) of one electron each. To compute the binding energy BE of the deuteron, we have the equation:

$$\mathrm{d} + \mathrm{BE} = \mathrm{p} + \mathrm{n}$$

where d is the deuteron, p is the proton, n is the neutron, and BE is the binding energy. As energy must be conserved, the energy content of the two sides must balance. Using the Einstein relation (15.7) we define the **mass defect** Δm of the deuteron to be the binding energy of the deuteron divided by the speed of light squared $\Delta m = \mathrm{BE}/c^2$. The mass defect may be computed as follows:

$$\begin{aligned}\Delta m &= m_p + m_n - m_d \\ &= (m_p + m_e) + m_n - (m_d + m_e) \\ &= 1.007825\,\mathrm{u} + 1.008665\,\mathrm{u} - 2.014102\,\mathrm{u} \\ &= 0.002388\,\mathrm{u}\end{aligned}$$

whereupon the binding energy of deuteron is computed to be

$$\mathrm{BE} = (\Delta m)c^2 = (0.002388\,\mathrm{u})(931.5\,\mathrm{MeV/u}) = 2.224\,\mathrm{MeV}$$

This much energy is required to separate the proton and neutron of a deuteron. Alternatively, if a proton and neutron are brought close together, they will release this much energy if they form a deuteron. (This is an example of a **fusion** reaction.) Note that the mass defect is not zero—the proton and neutron combined weigh more than the deuteron.

17. Relativistic Kinematics

17.1 Next we turn to an example of how one goes about computing something with the help of the relativistic formulae we have developed. The example clearly illustrates the interchangeability of mass and energy. [30]

17.2 Example. In 1955 G. Segré and his colleagues at the University of California at Berkeley produced **antiprotons** (the antiparticle of the proton, with the same mass but opposite charge, plus a few other opposite properties) by bombarding a stationary target containing hydrogen (protons) with a beam of protons. Various conservation laws (conservation of charge, baryon number, *etc.*) require that if an antiproton is created, a proton must be created as well. So, after the collision the incident and target protons remain, but we have created a proton-antiproton pair. The question is: *What is the minimum kinetic energy of an incident proton that can cause pair production?* This minimum kinetic energy is called the **threshold energy**.

As a first (but incorrect) guess, you might suppose that the minimum kinetic energy for pair production is achieved when **all** the incident kinetic energy is converted entirely to mass, leaving all four final particles (3 protons and one antiproton) at rest (see Figure 23). But a moment's thought reveals that this is impossible. Such a situation does not satisfy conservation of momentum.

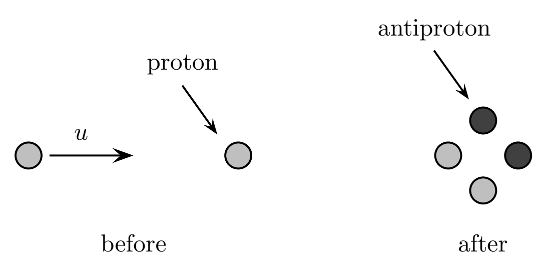


Figure 23. An impossible reaction in the laboratory frame.

[30] Adapted from E.F. Taylor and J.A. Wheeler, *Spacetime Physics* (W.H. Freeman, San Francisco, 1966).

On the other hand, there **is** a frame of reference in which all four particles end up at rest: it is called the **center-of-momentum** (or **c.o.m.**) **frame**, namely the frame in which the total momentum of the system is zero (see Figure 24). In this frame it is clear that the total energy of the incoming protons is a minimum when the four final particles remain together **and at rest** after the collision. If they were to move off with some kinetic energy, this energy would have to come from the incoming energies of two protons.

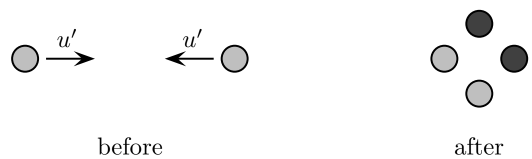


Figure 24. Threshold proton-antiproton pair production in the center-of-momentum frame.

The claim is that this situation, in which the four final particles remain together and at rest in the center-of-momentum frame after the collision, is the one that achieves the maximum conversion of kinetic energy to rest mass. It follows that, when we again consider the problem from the point of view of the laboratory frame, the threshold kinetic energy will be the one that results in all four final particles moving off as a group with the same final velocity (see Figure 25).

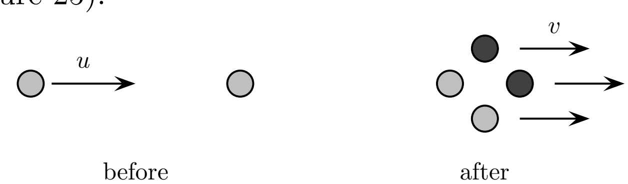


Figure 25. Threshold proton-antiproton pair production in the laboratory frame.

We may use this information to solve the problem. Before we do so, however, we adopt the convention explained in Section 10.11. That is, we agree to choose our units so that $c = 1$. We may then restore the factors of c at the end of the day. We have, from the law of conservation of momentum:

$$p = 4p' \tag{17.1}$$

and from the law of conservation of energy:

$$E + m = 4E' \tag{17.2}$$

where p and E are the initial momentum and energy of the incident proton, respectively, m is the rest energy of the target proton (remember, $c = 1$!), and p' and E' are the final momentum and energy of any one of the four particles.

We may rewrite (17.2) as

$$T + 2m = 4E' \tag{17.3}$$

where T is the threshold kinetic energy we seek. The invariant mass equation (15.15) yields

$$(4E')^2 - (4p')^2 = (4m)^2 \tag{17.4}$$

Inserting (17.1) and (17.3) into (17.4) we get

$$(T + 2m)^2 - p^2 = 16m^2 \tag{17.5}$$

Using (15.15) once more (this time for unprimed quantities) to substitute for p in (17.5) we get

$$(T + 2m)^2 - [(T + m)^2 - m^2] = 16m^2 \tag{17.6}$$

Expanding, we have

$$T^2 + 4Tm + 4m^2 - [T^2 + 2Tm] = 16m^2 \tag{17.7}$$

which becomes

$$2Tm = 12m^2 \tag{17.8}$$

or

$$T = 6m \tag{17.9}$$

Putting back the factors of c is easy; dimensional considerations dictate that

$$T = 6mc^2 \tag{17.10}$$

As the rest energy of a proton is about 1 GeV, we see that the threshold kinetic energy is about 6 GeV.

17.3 It is an amusing exercise to solve the problem in the c.o.m. frame then transform back to the laboratory frame. In the c.o.m. frame the proton entering from the left has momentum s, say, while the one on the right has momentum $-s$. They both have the same energy, say e. After the collision, the final momentum is, of course, zero, while the final energy is simply the rest energy of 3 protons and 1 antiproton. So the conservation laws of momentum and energy become

$$s - s = 0 \tag{17.11}$$

and

$$2e = 4m \tag{17.12}$$

To finish the problem, we must transform back into the laboratory frame to find the energy E of the incident proton. For the incoming proton we have

$$\begin{pmatrix} E \\ p \end{pmatrix} = \begin{pmatrix} \gamma_R & \beta_R\gamma_R \\ \beta_R\gamma_R & \gamma_R \end{pmatrix} \begin{pmatrix} e \\ s \end{pmatrix} \tag{17.13}$$

from which we conclude

$$E = \gamma_R e + \beta_R\gamma_R s \tag{17.14}$$

$$p = \beta_R\gamma_R e + \gamma_R s \tag{17.15}$$

while for the stationary proton we have

$$\begin{pmatrix} m \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_R & \beta_R\gamma_R \\ \beta_R\gamma_R & \gamma_R \end{pmatrix} \begin{pmatrix} e \\ -s \end{pmatrix} \tag{17.16}$$

from which we derive

$$m = \gamma_R e - \beta_R\gamma_R s \tag{17.17}$$

$$0 = \beta_R\gamma_R e - \gamma_R s \tag{17.18}$$

From (17.18) we conclude

$$\beta_R = \frac{s}{e} \tag{17.19}$$

Putting this into the invariant mass equation we get

$$\begin{aligned} m^2 &= e^2 - s^2 \\ &= e^2 - \beta_R^2 e^2 \\ &= e^2(1 - \beta_R^2) \end{aligned} \tag{17.20}$$

from which we conclude, with the help of (17.12)

$$\gamma_R = \frac{e}{m} = 2 \tag{17.21}$$

Adding (17.14) and (17.17) and using (17.21) gives

$$E + m = 2\gamma_R e = 4e = 8m \Longrightarrow E = 7m \tag{17.22}$$

whence we conclude

$$T = E - m = 6m \tag{17.23}$$

as before.

17.4 At the risk of belaboring the point, let us examine the results of this problem a little more closely. Observe from (17.3) and (17.9) that the final energy of each particle after the collision in the laboratory frame is

$$E' = 2m \tag{17.24}$$

This means that, of the $6m$ of kinetic energy of the incoming proton, $2m$ goes toward producing the proton-antiproton pair, and $4m$ goes toward the kinetic energy of the products (m each). In effect, most of the incoming energy is wasted (as the objective is to produce new particles, not simply fast ones). Had we done this experiment in a proton-proton collider, we would have only had to give the incoming particles a total of $2m$ of kinetic energy (m each). [31] This is why colliding beam accelerators are superior to fixed target accelerators.

[31] Note that this entire discussion ignores the electrostatic interactions of the particles. This is justified because the energy required to get the protons within 'striking distance' (around a fermi (10^{-15} m)) is on the order of an MeV, and a kinetic energy of m is around a GeV, which is a factor of 1000 higher.

Appendix A. Linear Transformations

***A*.1** A **vector space** V (over the real numbers) is a collection of elements called **vectors** satisfying the property that, if v and w are vectors, so is $av + bw$, for any two real numbers a and b. In particular, this means that arbitrary scalar multiples of vectors are vectors, and sums of vectors are vectors. You should be quite familiar with the concept of a vector space, as we have been modeling three dimensional space by the vector space $\mathbb{R}^3$. In this Appendix, however, we will restrict ourselves to the plane $\mathbb{R}^2$.

***A*.2** We can put coordinates on $\mathbb{R}^2$ as usual, in which case vectors are denoted by pairs (x, y) of real numbers. The rules of scalar multiplication and vector addition are, of course,

$$a \cdot (x, y) = (ax, ay)$$

and

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

respectively. With these rules, $\mathbb{R}^2$ satisfies all the properties of a vector space.

***A*.3** We may put different coordinates (x', y') on $\mathbb{R}^2$, with analogous rules of scalar multiplication and vector addition. The natural question that arises is then, what is the form of the transformation relating the coordinates? In general, we will have $x' = f(x, y)$ and $y' = g(x, y)$, where f and g are arbitrary functions.

***A*.4** At this point we may choose to view the transformation in two ways: **passively** or **actively**. According to the passive view, the points of $\mathbb{R}^2$ are not transformed. Instead, it is the assignment of coordinates to the points that is changed. On the other hand, according to the active point of view, the assignment of coordinates to the locations in the plane is fixed, but the points themselves are moved about. According to this view, the point whose coordinates were (x, y) is moved to the point whose coordinates are (x', y'). The passive and active viewpoints are just that, namely points of view. They are completely equivalent interpretations of the same transformation. But each one has certain advantages.

***A*.5** If we adopt the active point of view of the transformation, then we may view it as a transformation T that moves the vector $u = (x, y)$ to the vector $u' = (x', y')$. That is, T

is a **map** from $\mathbb{R}^2$ to itself that takes u to u'. We may write

$$\boxed{u' = Tu} \tag{A.1}$$

to denote this transformation.

***A*.6** We are finally in a position to define what we mean by a **linear transformation**. We say T is a linear transformation if

$$T(pu + qv) = pTu + qTv = pu' + qv' \tag{A.2}$$

for any scalars p and q and vectors u and v. That is, the image of a scalar multiple of a vector is the scalar multiple of the image, and the image of the sum of two vectors is the sum of the images. Speaking mathematically for a moment we say that a linear transformation is one that preserves the linear structure of the vector spaces between which it acts.

***A*.7** What are the implications of linearity of T for the functions f and g defining the transformation? Well, not surprisingly, they must be **linear**. That is, we must have $f(x, y) = ax + by$ and $g(x, y) = cx + dy$ for some scalars a, b, c, and d. Only functions of this form give rise to linear transformations. Let us check this with a simple example (the general proof follows similarly, but it is more cumbersome). Suppose $f(x, y) = 3x + 2y$ and $g(x, y) = 0$. Let $u = (1, 2)$ and $v = (2, 1)$. Then $u + v = (3, 3)$ and $T(u + v) = T(3, 3) = (15, 0)$. But $Tu = (7, 0)$ and $Tv = (8, 0)$, so we have $T(u + v) = Tu + Tv$, as required. On the other hand, suppose that $f(x, y) = x^2y$ and $g(x, y) = 0$. Then $T(u + v) = T(3, 3) = (27, 0)$. But $Tu = (2, 0)$ and $Tv = (4, 0)$ so $T(u + v) \neq Tu + Tv$. In other words, the functions f and g are **nonlinear**, so the transformation T is nonlinear.

Appendix B. Matrix Algebra

***B*.1** A **matrix** is an $n \times m$ array of numbers, where n denotes the number of rows and m the number of columns. Typically, we denote the matrix by capital Latin letters. For example,

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$$

is a 2×2 matrix. Similarly,

$$B = \begin{pmatrix} 0.7 & -4.2 & 3.3 \\ 0 & 1.1 & 4.0 \end{pmatrix}$$

is a 2×3 matrix. We shall only be concerned with the case in which $n = m$, in which case we call the matrices **square**. Indeed, in these notes we shall, for the most part, restrict ourselves to the case in which $n = m = 2$.

***B*.2** We define the **components** of a matrix A to be the numbers in the array. We index the components by row and column, and write A_{ij} to denote the entry in the i^{th} row from the top and j^{th} column from the left. We call A_{ij} the ij^{th} entry. In the example above, $A_{11} = 1$, $A_{12} = 4$, $A_{21} = 2$, and $A_{22} = 3$. For a square matrix, we define the **diagonal** to be the set of entries of the form A_{ii} for all i. In our example, the diagonal of A is $\mathrm{diag}(A) = (1, 3)$.

***B*.3** We can make the set of $n \times n$ matrices into an **algebra**, which is just a fancy way of saying that we can define addition and multiplication of matrices in such a way that (almost all of) the familiar rules of algebra hold. We do this as follows. First, let A and B be two $n \times n$ matrices. Then we define the **sum** $C = A + B$ of the two matrices as follows:

$$C_{ij} = A_{ij} + B_{ij} \tag{B.1}$$

That is, the ij^{th} entry of the sum is the sum of the ij^{th} entries of A and B. We say that matrices add **component-wise**. For example,

$$\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} + \begin{pmatrix} 0 & -2 \\ 5 & -4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 7 & -1 \end{pmatrix}$$

***B*.4** Next, we define **scalar multiplication** (that is, multiplication of a matrix by a number) in a similar fashion:

$$(aC)_{ij} = aC_{ij} \tag{B.2}$$

That is, the ij^{th} entry of the matrix aC is just the matrix obtained from C by multiplying each entry by the number a. For example,

$$3 \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 12 \\ 6 & 9 \end{pmatrix}$$

***B*.5** Finally, we define **matrix multiplication** as follows

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \tag{B.3}$$

That is, the ij^{th} entry of the product matrix AB is given by multiplying the entries of the i^{th} row of A by the entries of the j^{th} column of B, and summing them. At first sight, this seems a rather odd thing to do, but as you learn more mathematics you will discover that it is, in fact, quite natural. Let us write out the rule explicitly for 2×2 matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

For example,

$$\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 0 & -2 \\ 5 & -4 \end{pmatrix} = \begin{pmatrix} 20 & -18 \\ 15 & -16 \end{pmatrix}$$

***B*.6** We then extend these rules by means of the usual distributive rules of arithmetic. For example, $a(A+B) := aA + aB$, and so on. It is worth noting that, although matrix multiplication (B.3) is **associative** [so that $A(BC) = (AB)C$], it is **not** in general **commutative** (that is, in general $AB \neq BA$).

***B*.7** The only thing we have not yet defined is how to divide matrices. This is a somewhat more involved subject, but for our purposes here we only need the following facts. Recall that the inverse of a nonzero number a is the number b such that $ab = 1$. If the inverse b exists (that is, if the original number is nonzero), we denote it by a^{-1}. Before we can do something similar for matrices we need to know what the analogue of 1 is. Well, as far as numbers are concerned, 1 is the only number with the property that $1a = a1 = a$ for all numbers a.

***B*.8** By analogy, the **unit or identity matrix**, usually denoted I, is the unique matrix that satisfies the property that $AI = IA = A$ for all matrices A. If you ponder the definition (B.3) for a while, you will see that the identity matrix is the one that has 1s along the diagonal, and 0s elsewhere. If we are talking about 2×2 matrices, the identity is

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

***B*.9** Given a matrix A, if there exists a matrix B such that $AB = BA = I$, then we say that A is invertible with inverse B, and we denote it by A^{-1}. If the inverse exists, it is unique. The condition for an inverse to exist is that the matrix have nonzero **determinant**, where the determinant of a 2×2 matrix is given by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

The determinant of higher order matrices is more complicated, but can be found in any book on linear algebra.

***B*.10** Finally, we define the notation used in the text. Instead of writing a point in a coordinate system as (x, y), we may view it as a **column vector** $\begin{pmatrix} x \\ y \end{pmatrix}$. Then a linear transformation T carrying vectors to vectors may be represented as a matrix (also denoted T) acting on the components of the vector according to the rules of matrix multiplication. For example, if $u = (1, 2)$ and

$$T = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$$

then Tu is the vector

$$\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 9 \\ 8 \end{pmatrix}$$

We write $Tu = v$, where $v = (9, 8)$.

Appendix C. Hyperbolic Functions

***C*.1** According to J. Ratcliffe, one of the earliest works on hyperbolic geometry was the 1770 treatise of Lambert entitled *Observations Trigonométriques.* [32] Hyperbolic geometry may be viewed as the geometry of the hyperbolic sphere, namely the sphere of unit imaginary radius in Lorentzian space. We take a more utilitarian point of view, and define hyperbolic functions as follows. First, recall the Taylor series expansions of the elementary functions about the point 0:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + ... \tag{C.1}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + ... \tag{C.2}$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... \tag{C.3}$$

If we substitute ix for x in (C.3) (where $i = \sqrt{-1}$) we get

$$\begin{aligned} e^{ix} &= 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + ... \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + ... + i(x - \frac{x^3}{3!} + \frac{x^5}{5!} + ...) \end{aligned}$$

[32] See J.Ratcliffe, *Foundations of Hyperbolic Manifolds* (Springer-Verlag, New York, 1994)

Comparing this with the expansions (C.1) and (C.2) we discover the famous Euler formula:

$$\boxed{e^{ix} = \cos x + i \sin x} \tag{C.4}$$

It follows from the Euler formula that

$$\sin x = \frac{1}{2i}(e^{ix} - e^{-ix}) \tag{C.5}$$

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}) \tag{C.6}$$

C.2 We now define hyperbolic sine and cosine (pronounced "sinch" and "kosh") to be

$$\sinh x = \frac{1}{2}(e^{x} - e^{-x}) \tag{C.7}$$

$$\cosh x = \frac{1}{2}(e^{x} + e^{-x}) \tag{C.8}$$

By comparing (C.7) and (C.8) with (C.5) and (C.6) we see that the hyperbolic functions are related to the trigonometric ones by the following relations:

$$\cosh x = \cos ix \tag{C.9}$$

$$\sinh x = -i \sin ix \tag{C.10}$$

Clearly, sinh must be an odd function of its argument (*i.e.*, $\sinh(-x) = -\sinh x$), while cosh must be even (*i.e.*, $\cosh(-x) = \cosh x$).

C.3 One can define other hyperbolic functions analogously. For example,

$$\tanh x = \frac{\sinh x}{\cosh x} \tag{C.11}$$

We shall need two properties of these hyperbolic functions, which you can easily prove from the definitions (C.7) and (C.8). First

$$\cosh^2 x - \sinh^2 x = 1 \tag{C.12}$$

This is just Equation (11.16) in the text. Second,

$$\tanh(x + y) = \frac{\tanh x + \tanh y}{1 + \tanh x \tanh y} \tag{C.13}$$

This is the hyperbolic analogue of the angle addition formula for tangent.

Appendix D. The Relativistic Doppler Effect

***D*.1** For completeness, I include here a derivation of the relativistic Doppler shift formula. Recall that the ordinary Doppler shift refers to the phenomenon in which the frequency of a sound wave gets shifted up or down in frequency depending upon the motion of the source and observer. Something similar happens for light waves, but there is one primary difference. In the case of sound waves, the Doppler shift depends upon the motion of the source and the motion of the observer. This follows because sound waves travel through air, and one's motion through the air matters for one's perception of the sound. On the other hand, the relativistic Doppler shift depends only upon the **relative** motion of the source and observer. This is because the speed of light is independent of the motion of the source.

***D*.2** Consider first the case in which a source of electromagnetic waves and a receiver approach one another. We analyze the problem from the point of view of the receiver, who therefore determines the rest frame for the problem. Suppose the source moves toward the receiver with a speed v. Further suppose that the receiver observes the source emit N waves in a time Δt. In the time it takes the lead wave to travel a distance $c\Delta t$, the source has moved forward a distance $v\Delta t$ (see Figure 26).

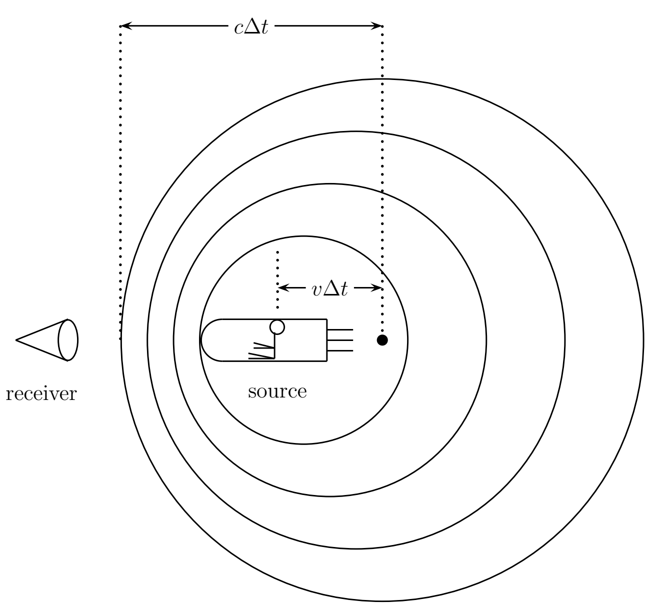


Figure 26. The relativistic Doppler effect.

***D*.3** The perceived wavelength λ is

$$\lambda = \frac{1}{N}[c\Delta t - v\Delta t] \tag{D.1}$$

whence the observed frequency is

$$\nu = \frac{c}{\lambda} = \frac{c}{c-v}\frac{N}{\Delta t} = \frac{1}{1-\beta}\frac{N}{\Delta t} \tag{D.2}$$

***D*.4** Now, let the frequency of the source be ν_0 (in its own restframe). It emits $N = \nu_0 \Delta t'$ waves, where $\Delta t'$ is the time between the emission of the first and last waves as measured by an observer moving with the source. (Both observers agree on the number of waves emitted, because they can count them.) As $\Delta x' = 0$ (the waves are all emitted from the same location in the source frame) it follows that

$$\Delta t = \gamma \Delta t' \tag{D.3}$$

so that

$$\begin{aligned} \nu &= \frac{1}{1-\beta} \frac{\nu_0 \Delta t'}{\Delta t} \\ &= \frac{1}{1-\beta} \frac{\nu_0}{\gamma} \\ &= \frac{\sqrt{1-\beta^2}}{1-\beta} \nu_0 \end{aligned}$$

We conclude that

$$\boxed{\nu = \nu_0 \sqrt{\frac{1+\beta}{1-\beta}}} \tag{D.4}$$

for the case in which the source and the receiver approach one another with relative speed β.

***D*.5** It follows immediately (merely by sending β to $-\beta$) that

$$\boxed{\nu = \nu_0 \sqrt{\frac{1-\beta}{1+\beta}}} \tag{D.5}$$

for the case in which the source and the receiver recede from one another with relative speed β.

18. Problems

Note: Unless stated otherwise, in the following problems the rocket moves in the positive x direction with speed β.

1. Time Dilation with Muons. In a given sample of muons, half will decay via the channel $\mu \to e^- + \bar{\nu}_e + \nu_\mu$ in 1.5 microseconds (measured in the laboratory frame that is at rest relative to the muons). Half of the remainder will decay in the next 1.5 microseconds, and so on. (We say that the muons have a *half-life* of 1.5 microseconds.)

a. Consider muons produced by the collision of cosmic rays with gas nuclei in the atmosphere at a height of 60 kilometers above the surface of the earth. The muons move vertically downward with a speed approaching (but not equal to!) that of light. Approximately how long will it take them to reach the Earth as measured by an observer at rest on the surface of the Earth? [Hint: You do not need any relativity for this one!]

b. If there were no time dilation, approximately what fraction of muons produced at a height of 60 kilometers would remain undecayed by the time they reached the surface?

c. It is found experimentally that the fraction of muons that survive the trip to the Earth's surface is approximately 1/8! Using this information, deduce the difference between the speed of light and the speed of the muons. That is, compute the quantity $1 - \beta$. [Note: You will have to use Taylor's theorem.]

2. The Headlight Effect. A flash of light is emitted at an angle ϕ' with respect to the x' axis of the rocket frame.

a. Show that the angle ϕ that the direction of this flash makes with respect to the x axis of the laboratory frame is given by the equation

$$cos\ \phi = \frac{cos\ \phi' + \beta}{1 + \beta\ cos\ \phi'}$$

b. Now consider a particle at rest in the rocket frame that emits light uniformly in all directions. Let $\beta = 0.9$. Consider the 50% of this light that goes into the *forward* hemisphere in the rocket frame. Show that in the laboratory frame this light is concentrated in a *narrow forward cone* whose axis lies in the direction of motion of the particle, and find the opening angle of the cone. This effect is called the *headlight effect*. It is observed in particle accelerators.

3. Symmetric Elastic Collision. A particle of mass m and kinetic energy T collides elastically with a particle of the same mass at rest. The two particles emerge with equal energies, so their velocity vectors make equal angles $\alpha/2$ with respect to the initial direction of motion. In Newtonian mechanics one can prove (see Halliday and Resnick Chapter 10) that the two particles always separate with a relative angle $\alpha = 90$ degrees. This is no longer true in relativistic mechanics.

a. Using only the relativistic laws of conservation of energy and momentum, show that

$$cos^2(\alpha/2) = \frac{T + 2mc^2}{T + 4mc^2}$$

b. Using the trigonometric identity

$$cos^2(\alpha/2) = \frac{1 + cos\ \alpha}{2}$$

show that

$$cos\ \alpha = \frac{T}{T + 4mc^2}$$

c. What is α for a Newtonian elastic collision at low velocity? What is α for an extreme relativistic collision with large T?

4. Photoproduction of a Positron-Electron Pair by a Single Photon. A gamma ray γ (high energy photon, zero rest mass) can carry an energy greater than the rest energy of an electron-positron pair. (Remember that a positron e^+ has the same mass as an electron e^- but opposite charge.) Nevertheless, the process

$$\gamma \longrightarrow e^- + e^+$$

cannot occur in the absence of other matter or radiation.

a. Prove that the above process is incompatible with the laws of conservation of momentum and energy as employed in the laboratory frame of reference. Treat the most general case, in which the paths of the alleged electron and positron do *not* make equal angles with the extended path of the incoming gamma ray.

b. Now repeat the demonstration–which then becomes much more impressive–in the center-of-momentum frame of the alleged pair (the frame of reference in which the total momentum of the resulting particles is zero). [Hint: Think of the time reversed process $e^- + e^+ \rightarrow \gamma$. The laws of physics are (for the most part) time-reversal invariant, so that, if the laws of physics were to permit this reaction they would permit the time-reversed reaction.]

c. In the presence of other matter, a gamma ray *can* produce an electron-positron pair. What is the threshold energy T_{th} at which a gamma ray becomes capable of bringing about the (often observed) process

$$\gamma + e^- \text{ (at rest)} \longrightarrow e^+ + 2e^-$$

The rest energies of the electron and positron are 0.511 MeV.

5. The Physicist and the Traffic Light. A physicist is arrested for going through a red light. In court he pleads that he approached the intersection at such a speed that the light looked green to him. The judge, who has taken undergraduate physics, changes the charge to speeding and fines the defendant one dollar for every mile per hour he exceeded the local speed limit of 20 mph. What is the fine? Take the wavelength of green light to be 5300 Ångstroms (1 Å$= 10^{-10}$ m), and that of red light to be 7000 Å.

6. Satellite Transmissions and the Doppler Effect. An Earth satellite, transmitting on a frequency of 40 MHz, passes directly over a radio receiving station at an altitude of 400 km and at a speed of 2.8×10^4 km/h. Plot the change in frequency, attributable to the Doppler effect, as a function of time, counting $t = 0$ as the instant the satellite is over the station. [Hint: The speed in the Doppler formula is not the actual speed of the satellite but its component in the direction of the station. Neglect the curvature of the Earth and of the satellite orbit.]