An Introduction to Special Relativity

Lambert E. Murray*

Department of Physical Science
Harding University, Searcy, AR 72149-0001

(Dated: August 31, 2012)

1. INERTIAL REFERENCE FRAMES AND THE PRINCIPLE OF RELATIVITY

1.1. Newton's Three Laws of Motion

A clear understanding of how objects move near the earth's surface begins with Newton's three laws of motion:

1. The Law of Inertia: an object moving in a straight line will continue to move along that same straight line at a constant rate of speed unless acted upon by some outside, unequal force. This law, in effect, defines the *natural* state of matter as one of *constant velocity*. Now:

   - Velocity is a *vector* with both magnitude *and* direction.
   - The *components* of this velocity vector depend upon the choice of coordinate system.
   - The *zero* velocity vector is also a possible *natural* state of an object.

   But how do we define constant velocity? Is there a reference frame *fixed* in space relative to which we can measure this constant velocity? Although we might have some difficulty in coming up with a precise way of defining a "fixed" coordinate system from which we can measure constant velocity, we *define* an *inertial reference frame* to be one in which the law of inertia is valid.

2. An object *changes* its natural state when acted upon by a net *external* force: the *rate of change* of the velocity is proportional to the size of the applied force and inversely proportional to the *inertial mass* of the object. This is usually written mathematically as

$$\overrightarrow{F} = m\overrightarrow{a} \tag{1}$$

   where $\overrightarrow{a} = d\vec{v}/dt$.

3. For every action there is an *equal* and *opposite* reaction. This law essentially states that forces of nature always occur in *pairs*.

*Electronic address: lmurray@harding.edu

1.2. Intertial Reference Frames and the Principle of Galilean Relativity

Let's look more closely at some of the implications of Newton's laws of motion. First we will *assume* that we have an idealized, fixed coordinate system (for the moment ignoring the problem of actually finding such a coordinate system), and that this coordinate system is a Cartesian coordinate system with orthogonal axes labeled x, y,and z. We will call this the **Home Frame**. Let a vector $\overrightarrow{R}$ locate an object relative to the origin of this fixed coordinate system. This vector can be expressed in terms of its components along the x, y, and z axes in the form:

$$\overrightarrow{R} = \widehat{i}x + \widehat{j}y + \widehat{k}z = (x, y, z) \tag{2}$$

Now, in general, the position of the object may change in time. We therefore *define* the velocity vector as the time rate of change of the position vector $\overrightarrow{R}$, or

$$\overrightarrow{V}(t) = \frac{d\overrightarrow{R}(t)}{dt} = \dot{\overrightarrow{R}}(t) = \widehat{i}\dot{x} + \widehat{j}\dot{y} + \widehat{k}\dot{z} = (\dot{x}, \dot{y}, \dot{z}) \tag{3}$$

where a dot above a quantity stands for the time derivative.

In addition, the velocity vector itself may be a function of time. It may change in magnitude, or in direction, or both. We *define* the acceleration vector as the time rate of change of the velocity vector $\overrightarrow{V}(t)$, or

$$\overrightarrow{A}(t) = \frac{d\overrightarrow{V}(t)}{dt} = \dot{\overrightarrow{V}}(t) = \widehat{i}\dot{v}_x + \widehat{j}\dot{v}_y + \widehat{k}\dot{v}_z = (\dot{v}_x, \dot{v}_y, \dot{v}_z) \tag{4}$$

Now, according to the second law, when the net external force acting on an object is zero, the acceleration of that object is also zero, so that the velocity of the object must be a constant. This is the *natural* state of the object.

Let's assume that an object is moving at constant velocity along the x-axis of a fixed coordinate system. We now construct a *new* coordinate system which we will designate as the x', y', z' coordinate system (the **Other Frame**) with axes parallel to the original x, y, z coordinate system and which moves along the $+x$ axis at the same speed as the object under consideration. In this new coordinate system, the object is *at rest*! Both of these coordinate systems are inertial, because the law of inertia holds in both systems: if there is no net external force acting on the object, the velocity of the object remains constant. Thus, *any coordinate system moving at constant velocity is an inertial reference system in which the laws of physics are the same.* This is a statement of the *principle of relativity*! To make this argument a little more concrete, consider the diagram shown in Figure 1. The location of the object relative to the fixed reference frame is given by x, y, z and the location of the object relative to the moving reference frame is given by x', y', z' where $x = x' + vt$, $y = y'$, and $z = z'$. Taking the time derivative of these equations gives

$$\begin{aligned} \dot{x} &= \dot{x}' + v \\ \dot{y} &= \dot{y}' \\ \dot{z} &= \dot{z}' \end{aligned} \tag{5}$$

from which we derive Galileo's law of velocity addition. This is the law we use when we determine the velocity of a speedboat relative to the shore if we know the velocity of the boat relative to the moving water and the velocity of the water relative to the shore. Similarly,

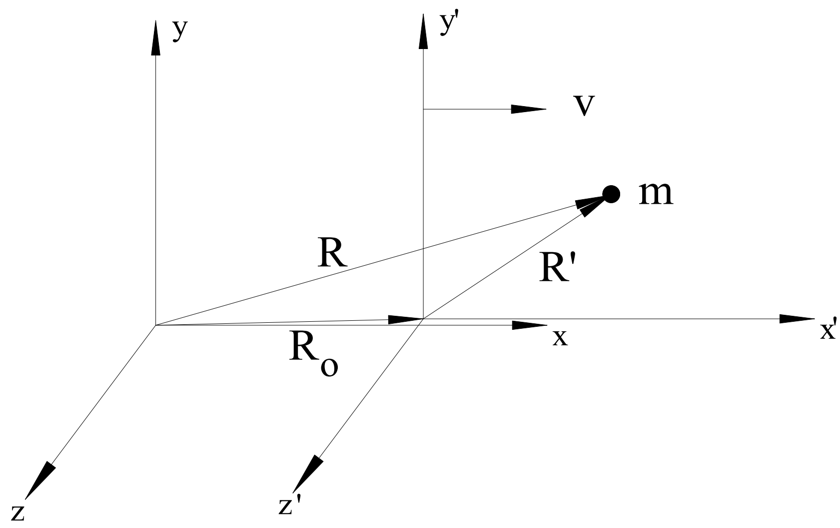

FIG. 1: The location of a particle as seen in two inertial reference frames, one moving with a speed v relative to the other.

these are the equations used when we wish to determine the velocity of a plane relative to the ground when we know the velocity of the wind and the velocity of the plane relative to the air. Notice that in deriving these equations we made the *reasonable* assumption that time intervals measured in the *moving* frame are equivalent to time intervals in the *rest* frame.

EXERCISE 1.1 *An airplane flies with speed c relative to still air from point A to point B and then returns. Compare the time, $t_{parallel}$, required for the round trip when the wind blows from A to B with speed v with the time, $t_{perpendicular}$, when the wind blows perpendicular to the line AB with speed v, (i.e., compute $t_{parallel}/t_{perpendicular}$).*

Now, since the moving reference frame is moving with constant velocity, if we take the time derivative of the velocity vectors in the last equation, we obtain

$$\begin{aligned} \ddot{x} &= \ddot{x}' \\ \ddot{y} &= \ddot{y}' \\ \ddot{z} &= \ddot{z}' \end{aligned} \tag{6}$$

or equivalently,

$$m\overrightarrow{A} = m\overrightarrow{A}' \tag{7}$$

which demonstrates that *Newton's laws of motion are identical in these two reference frames.* Thus, we conclude that the laws of physics should be the same in any two inertial reference frames (i.e., in any two reference frames moving with a constant velocity relative to one another). This means that there is no way to tell, based upon physical measurements, if a coordinate system is moving or not. We may be able to tell if something is moving relative to us, but we cannot determine *absolute* motion.

1.3. Maxwell's Equations, the Velocity of Light, and the Principle of Special Relativity

In the last section we showed that two reference frames moving with constant velocity relative to each other are equivalent reference frames as far as the laws of physics are concerned - or at least as far as Newton's second law is concerned, since the acceleration of an object is the same in both reference frames. One of the assumptions made in that section is that the velocity of an object measured in the Home Frame is the velocity of that same object measured in the Other Frame plus the relative velocity of the two reference frames. The equivalence of two inertial frames in which the velocities are additive is called the *principle of Galilean relativity.* As we also pointed out, *the principle of Galilean relativity depends upon the assumption that time intervals are equivalent in all inertial reference frames.*

In 1873 James Clerk Maxwell published a set of equations which summarized our understanding of electric and magnetic fields, and their interrelationship. One of the important consequences of these equations is that they predict the existance of electromagnetic waves. In its simplest form, the electromagnetic wave equation is given by

$$\frac{\partial^2 E_x}{\partial z^2} - \epsilon_o \mu_o \frac{\partial^2 E_x}{\partial t^2} = 0 \tag{8}$$

where ϵ_o and μ_o are constants known as the permittivity constant and the permeability constant, respectively. The solution to this differential equation is of the form

$$E_x(t) = E_{xo} \cos(kz - \omega t) \tag{9}$$

where the velocity of the traveling wave is given by

$$v = \frac{\omega}{k} = \frac{1}{\sqrt{\epsilon_o \mu_o}} \tag{10}$$

This means that the velocity of an electromagnetic wave measured in the laboratory depends only upon the two electrostatic constants ϵ_o and μ_o. If Galilean relativity is correct, the *velocity* of light should depend upon the velocity of the reference frame in which the measurements are obtained. But this would imply that the electrostatic constants would also depend upon the velocity of the reference frame in which electrostatic measurements were made. If this were true, the laws of electromagnetism would *not* be the same in all inertial reference frames.

A number of attempts have been made to measure a *variation* in the speed of light which would be consistent with Galilean relativity, but these attempts have all failed. Experimental evidence so far indicates that the speed of light is a constant in *all* inertial reference frames. This would seem to indicate that the principle of relativity should be extended to include *all* of physics (not just mechanics).

We therefore state, as a fundamental postulate of physics, that *the laws of physics are equivalent in all* ***inertial*** *reference frames.* This is a statement of what we call *the principle of* ***special*** *relativity*, as first presented by Albert Einstein. [Einstein later extended this principle to accelerating reference frames (known as *the principle of* ***general*** *relativity)* which is beyond the scope of this text.]

The conclusion that *the speed of light is the same in all inertial reference frames* means, of course, that the Galilean velocity transform is incorrect. Thus, we must re-examine the laws of mechanics from which the Galilean velocity transform is derived. As was also pointed out earlier, the Galilean transformation is based upon the *assumption* that *time intervals are equivalent in all inertial reference frames.* As we examine the consequences of the special theory of relativity (that the speed of light is the same in all reference systems), we will be lead to some very interesting and unexpected conclusions regarding time intervals and distance measurements in inertial reference frames.

2. SPECIAL RELATIVITY: A NEW LOOK AT TIME AND DISTANCE

2.1. Time Intervals in Different Inertial Frames

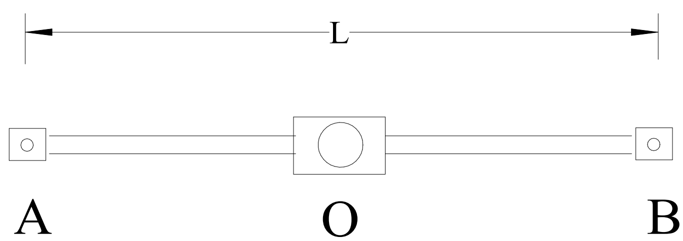

FIG. 2: A light source O is located equidistant from two detectors A and B. When light from source O reaches a detector, it sends out a green pulse of light.

To get a clear picture of the implications of the special theory of relativity we will examine a very special situation. Consider a rigid rod of length L with a light source located at the mid-point of the rod as shown in Figure 2. Let's assume this light source emits white light in all directions when triggered. At each end of the rod we place a special light sensitive detector. The instant either detector detects a light pulse, that detector will emit a light flash (we will assume that this light flash is green). At some instant of time t_1 the light source is triggered. The light from the source radiates spherically outward in all directions and the light pulse reaches each end of the rod *simultaneously*, i.e.,

$$t_A = t_B = \frac{L/2}{c}, \tag{11}$$

where c is the speed of light.

In order to precisely describe the situation in a totally unambiguous manner, let us carefully talk about the different *events* described in this situation. We will use the term *event* to describe *the location and the time* that a certain thing occurs. We can then record these events on a space-time diagram, like the one shown in Figure 3. On this diagram the initial light pulse at O and the subsequent flashes from the detectors at points A and B can be expressed in terms of *when* and *where* the event occurred using a coupled pair, e.g., (t'_A, x'_A). Thus, the *event* O (a flash of light emitted from the source) occurs at the origin of the coordinate system $x' = 0$ at time $t'_0 = 0$, giving an event $(t'_0 = 0, x'_0 = 0)$. As indicated on this diagram the location of the source O and the detectors A and B remain *fixed* (i.e., constant x' values) as time t' increases, creating vertical *world lines* for the source and detectors. A world line is the line traced out by an object in a space-time diagram.

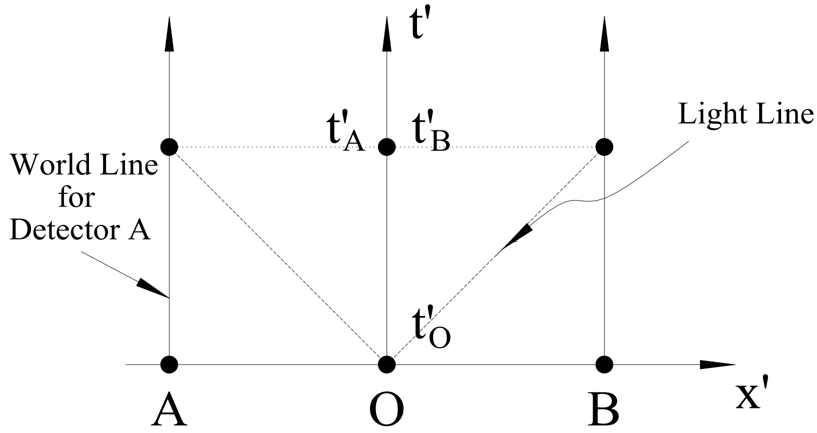

FIG. 3: A space-time diagram illustrating the simultaneous detection of light from source O by the detectors A and B. This diagram represents the events as seen in the reference frame of the rod.

Just as in introductory physics, a straight line represents constant velocity in one dimension, a positive slope corresponding to positive velocity, and a negative slope corresponding to negative velocity. However, the slope of the line in a space-time diagram is *inversely* related to the velocity, since the time and space axes are reversed from the way they are commonly done in classical mechanics.

Thus, in the figure, the light pulse which originates at point O at time $t'_0 = 0$ propogates in the plus and minus x' directions until it reaches the detectors A and B simultaneously at time $t'_A = t'_B$. On this diagram the *light line* (the world line for light) has a slope of 1 since the *distance* scale is typically calibrated in light-seconds while the time scale is calibrated in seconds. The world line for the source and detectors have a slope of infinity ($slope = 1/v$) since they are not moving in this coordinate system.

Now, let's examine this same set of events as seen in *another* inertial reference frame. Although you may have assumed that our experimental apparatus (the rod with light source and detectors) was originally *fixed* in space, that is not necessarily the case. Remember that all inertial reference frames are equivalent, so that the rod could be moving at some constant speed v. We will call the reference frame in which the rod is at rest the S' frame. Nothing in our previous discussion would be altered.

We now consider the motion of the rod as seen in another inertial reference frame (the S frame) in which the rod appears to be moving at constant speed in the $+x$-direction, parallel to the x axis as shown in Figure 4. We again want to represent the chain of events using a space-time diagram. In this case the location of the light source (the origin of the S' frame) and the detectors A and B are all moving in the $+x$ direction with a speed v. We can represent this on the space-time diagram as shown in Figure 5. The world lines of detectors A and B and of the source O are drawn as slanted lines, each with the same slope of $1/v$. Again, the space-time diagram is scaled so that one light-second of distance is the same as one second of time so that a light beam *which, according to the special theory of relativity, must travel at the same speed in all inertial frames* still has unit slope. We can see, using this diagram, that in this frame of reference, light will be received at detector A *before* it is received at detector B. Thus, according to the principle of special relativity *events which are*

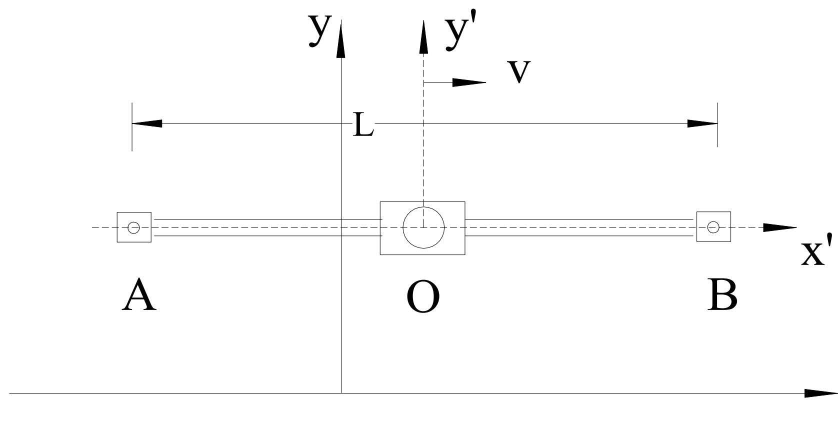

FIG. 4: The light-rod as seen in a reference frame where the rod moves by at constant velocity v.

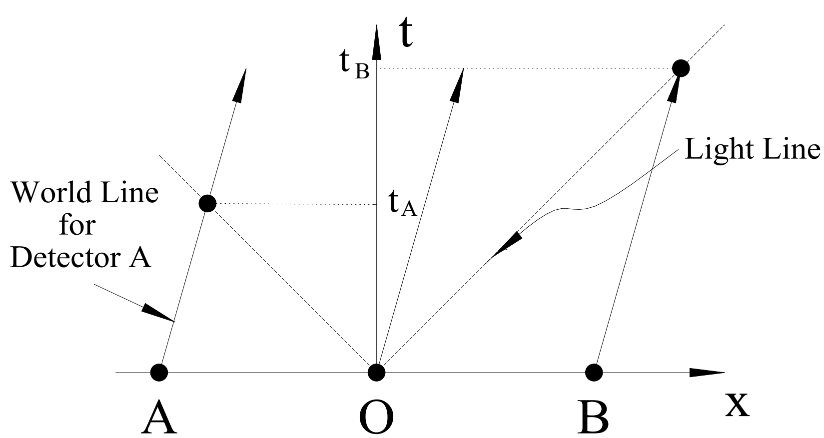

FIG. 5: A space-time diagram illustrating the detection of light from source O by the detectors A and B. This diagram represents the events as seen in a reference frame where the rod is moving parallel to the x-direction with a speed β.

simultaneous in one inertial reference frame are ***not*** *necessarily simultaneous in another.* In this particular example, a light flash is emitted from the source at time $t = t' = 0$ as the origins of the two reference frames coincide. In the frame which is moving with the light rod (the S' frame), the two detectors A and B detect the light at the same time, so that $t'_A = t'_B$. In the S frame, the light is detected *first* by the *trailing* detector (detector A) and then later by the leading detector (detector B), so that $t'_A < t'_B$. Therefore, if this light flash were used to synchronize clocks located at points A and B, an observer in the S' frame would claim that clocks A and B were simultaneously set to the correct time $t' = L/2c$. An observer in the S frame, however, would claim that clock A was set *earlier* than clock B.

2.2. Length Measurements in Different Inertial Frames

The fact that observers in two different inertial reference frames are unable to agree on whether or not events are simultaneous has some interesting implications on the measurement of the length of an object measured in these two reference frames. To understand how this occurs, consider a fish swimming in a fish bowl (see Figure 6). How can we determine

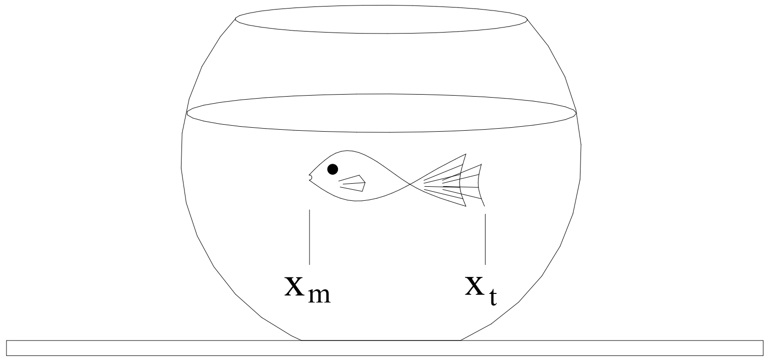

FIG. 6: Locating the ends of a moving fish in a fish bowl.

the length of the fish without removing the fish from the bowl? We first detemine the location of the mouth of the fish x_{mouth} at time t_{mouth} and then determine the location of the tail of the fish x_{tail} at some time t_{tail}. Notice the careful use of the concept of an *event*- we must specify both the location *and* the time to accurately describe the measurement process. If the fish is moving forward, and you determine the location of the mouth at one instant of time and determine the location of the tail at a *later* time, you will measure the fish to be *shorter* than it actually is. If, on the other hand, you determine the location of the tail at one instant of time and determine the location of the mouth at a *later* time, you will measure the fish to be *longer* than it actually is.Thus, the *length* L of the fish can be expressed by the equation

$$L = x_{mouth} - x_{tail} \tag{12}$$

only if the two locations x_{mouth} and x_{tail} are measured *simultaneously*! But, as was demonstrated above, two events which are simultaneous in *one* inertial frame of reference may *not* be simultaneous in another. This implies that the *length* of an object measured in one inertial reference frame may be *different* from the length of that same object measured in another inertial frame which is moving relative to the first!

This means that our definition of the *length* of an object must be re-examined! Since the length of an object depends upon the reference frame of the observer, we define the *proper length* of an object as that length *measured in a frame of reference which is at rest relative to the object*. This would be like sitting on the fish and using a tape measure to actually measure the length of the fish. In the rest frame of the fish, the location of the mouth and tail *does not change in time*, so that the location of either end of the fish can be determined *at any time*. Thus, when we speak of the length of a table in the laboratory, we are typically speaking of the *proper* length of the table.

3. RELATING TIME AND DISTANCE MEASUREMENTS IN DIFFERENT INERTIAL FRAMES: THE METRIC EQUATION

3.1. Relating *Time* Measurements in Different Inertial Frames

To determine quantitatively how to relate time measurements made in two inertial reference frames moving with a speed v relative to one another, we will examine a hypothetical experiment. In this experiment a light source and detector with a clock mechanism are attached to a rigid rod. A mirror is attached to this rod a distance L away from the light source (see Figure 7). When the clock reads $t' = 0$, the source sends out a pulse of light

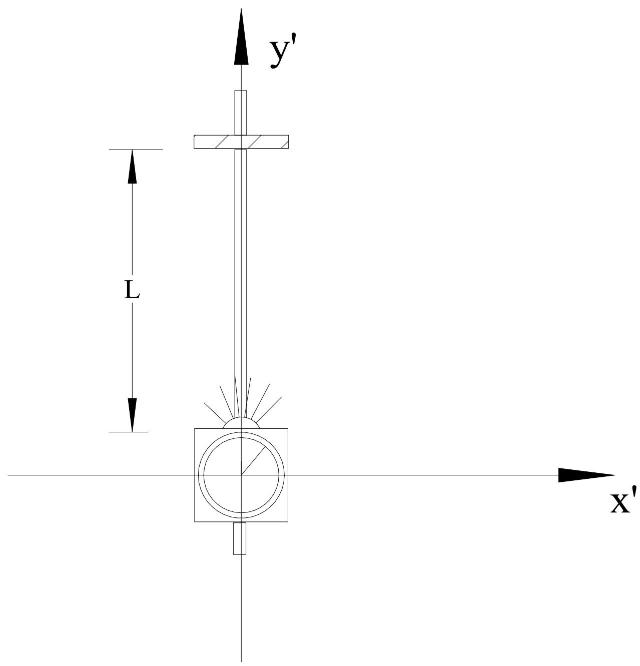

FIG. 7: A light-clock apparatus used to measure time intevals in different inertial reference frames.

(event A). This light pulse travels to the mirror, reflects off the mirror and returns to the detector which stops the clock (event B). Since the distance L can be precisely known (we are in the rest frame of the clock and mirror and can measure this distance at any time to any precision we desire), and since the speed of light in all reference frames is known to be c, we can determine the time interval $\Delta t'$ exactly. This time interval is given by

$$\Delta t' = 2L/c \tag{13}$$

Now, consider a reference frame through which this same light-clock mechanism is observed to move with a constant velocity v in a direction *perpendicular* to the rod supporting the mirror as shown in Figure 8. We will call this the HOME frame. In this reference frame event A (the initial pulse) and event B (the detection of that pulse) do *not* occur at the same *place*. An observer in the HOME frame, however, can still record the *events* as they occur, being careful to designate the *time and location* of each event. The time interval Δt between the two events as measured by an observer in the HOME frame is *different* from the time interval $\Delta t'$ measured by an observer *at rest* with respect to the mechanism. To

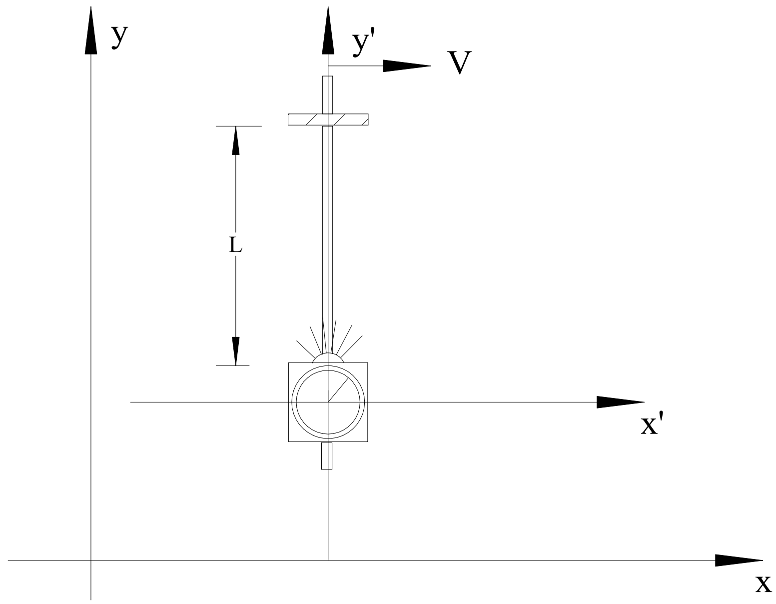

FIG. 8: The light-clock mechanism moving through another reference frame at constant velocity.

determine how the time interval Δt is related to the time interval $\Delta t'$, consider the diagram shown in Figure 9. The *light-clock* moves a distance

$$\Delta x = v\Delta t \tag{14}$$

as the light pulse travels from the source, reflects from the mirror, and then travels back to the detector (i.e., the distance between events A and B as seen in the HOME frame is the distance Δx). During this same time period, the *light pulse* moves a total distance of $2d$, where

$$d = \sqrt{L^2 + (\Delta x/2)^2} \tag{15}$$

is the diagonal distance from the original position of the light source (event A) to the location of the mirror when the light reaches that point and is reflected. Thus, the *total* time period between events A and B as measured in the HOME frame of reference is given by

$$\Delta t = \frac{2d}{c} = \frac{2\sqrt{L^2 + (\Delta x/2)^2}}{c} \tag{16}$$

$$= \frac{\sqrt{(2L)^2 + (\Delta x)^2}}{c} \tag{17}$$

$$= \sqrt{\left(\frac{2L}{c}\right)^2 + \left(\frac{\Delta x}{c}\right)^2} \tag{18}$$

But $2L/c$ is the time interval $\Delta t'$ between events A and B as measured in the reference frame attached to the moving light-clock mechanism. Thus, the time interval Δt, measured in the HOME frame, can be related to the time interval $\Delta t'$ measured in the *moving* S'

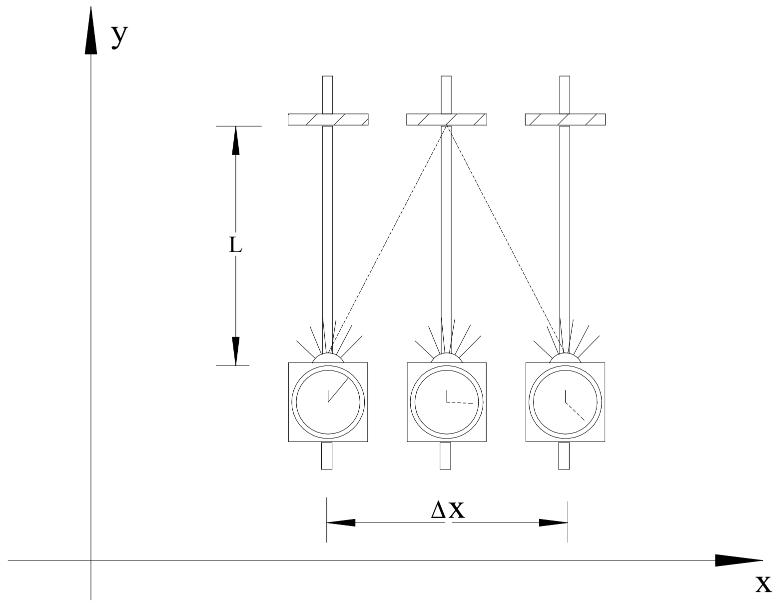

FIG. 9: The light-clock mechanism moves a distance $\Delta x = v\Delta t$ through this reference frame as the light pulse travels from the source to the detector.

frame by the equation

$$\Delta t = \sqrt{\Delta t'^2 + \left(\frac{\Delta x}{c}\right)^2} = \sqrt{\Delta t'^2 + \left(\frac{v\Delta t}{c}\right)^2} \tag{19}$$

Now, solving for the time interval $\Delta t'$, we obtain

$$\Delta t'^2 = \Delta t^2 - \left(\frac{\Delta x}{c}\right)^2 = \Delta t^2 - \left(\frac{v\Delta t}{c}\right)^2 \tag{20}$$

We can express the time interval $\Delta t'$ measured in the frame of reference of the moving clock to the time interval Δt measured in the HOME frame in two distinct ways. One is a simplified version of what we will later call the *metric* equation:

$$c^2\Delta t'^2 = c^2\Delta t^2 - \Delta x^2 \tag{21}$$

This equation relates the difference in time intervals measured by clocks in the two different reference frames to the *distance* between the two events as measured in one of the two reference frames. (We will show later that the quantity $(c\Delta t)^2 - \Delta x^2 = (c\Delta t')^2 - \Delta x'^2$ is an invariant quantity.)

Another useful way of expressing the relationship between the two time intervals is obtained by factoring out the time interval Δt on the right-hand side of Equation (20) to obtain

$$\Delta t' = \Delta t\sqrt{1 - \left(\frac{v}{c}\right)^2} \tag{22}$$

In this equation the difference in time intervals measured by clocks in the two different reference frames is related to the relative *velocity* of the two reference frames.

Just as the measured *length* of the fish in the fishbowl depends upon our reference frame, the *time interval* between any two events also depends upon our reference frame. And just as we define the *proper length* of an object as the length measured in the rest frame of the object, we define the *proper time interval* $\Delta\tau_o$ between two events as *that time interval measured between two events which occur at the same place.* (This is often stated as *the time interval that can be measured by the same clock.*) In the example we have just discussed, the time interval $\Delta t'$ must be the proper time interval, since events A and B both occur at the same place in the S' frame of reference. Thus, the proper time interval for this situation can be expressed as

$$\Delta\tau_o^2 = \Delta t^2 - \left(\frac{v\Delta t}{c}\right)^2 = \Delta t^2 - \left(\frac{\Delta x}{c}\right)^2 = \Delta t^2 \left[1 - \left(\frac{v}{c}\right)^2\right] \tag{23}$$

You should realize that the proper time interval, as defined in this last equation, is measured by a *single* clock in the rest frame of the light-clock mechanism and, therefore, *cannot be a function of the speed of the light-clock mechanism relative to any other reference frame.* You should also realize that in our derivation of the equations relating the proper time with the time interval measured in the HOME frame we have assumed that the light-clock mechanism is moving at a constant speed. This means that our light-clock mechanism is an *inertial* clock. Thus, the proper time interval that we have defined above is an *inertial* proper time interval. Since this inertial proper time interval is an invariant (i.e., it cannot be a function of the velocity of the reference frame) we give it a special name - the spacetime interval.

We should point out that it is possible to have a single clock that moves from one point to another recording the time interval between two separate events which does *not* move with constant velocity. The time interval measured by such a single clock would be a *proper* time interval, since the time interval is measured by a single clock, but it is not an *inertial* proper time interval. For this reason, we us the zero-subscript to distinguish an *inertial* proper time interval from a *non-inertial* proper time interval.

Equation (23) relates the inertial proper time interval $\Delta\tau_o$ measured between two separate events A and B *by a single clock* to the time interval Δt measured at two different locations within the HOME frame. If we assume that there is a clock located at each point within the HOME frame so that a given event can be precisely recorded, we realize that the time interval between the two events A and B which occur at two different locations must be measured by two different clocks. Since these two clocks are located at different coordinates within the HOME frame, the time interval Δt measured by these two clocks is called a *coordinate* time interval. Thus, equation (23) relates the inertial proper time interval between events A and B to the coordinate time interval between events A and B. It should be obvious that although the inertial proper time interval is an invariant, the *coordinate* time interval *does* depend upon the relative speeds of the two different frames of reference, and that the coordinate time interval depends upon the *distance between the two events as measured in the corresponding frame of reference.* Since the distance between any two observed events depends upon the relative speeds of the different reference frames, *the coordinate time interval must necessarily be different in different frames of reference.* Thus, *there are a large number of coordinate time intervals which could be measured for the same two events*, *but only one spacetime interval*!

Now when $v = 0$ the distance moved by the light-clock mechanism is zero, so that the coordinate time interval and the inertial proper time interval are the same. In fact, for speeds where $v/c << 1$ the two time intervals are *essentially* the same, consistent with the Galilean velocity transformations of classical physics. As v becomes larger, however, $\sqrt{1-(v/c)^2}$ becomes smaller and smaller. Thus, the inertial proper time interval between two events (the time interval measured by a single clock, moving at constant speed) will always be *shorter* than the time interval measured in any *other* reference frame. This phenomena is called *time dilation*. It is interesting to note here that as $v/c \to 1$, $\Delta\tau_o \to 0$ for *any* Δt. This means that for a photon, which travels at the speed of light, the time interval is *zero* between any two events at which the photon is present, i.e.,*a photon's clock does not tick!*

In our discussion thus far, we have assumed implicitly that the speed of the moving frame of reference is *less* than the speed of light. We will come back to this point later and ask if it is possible for the moving reference frame to move *faster* than the speed of light.

EXERCISE 3.1 *Use MatLab to plot the coordinate time interval as a function of the relative speed ($\beta = v/c$) of a moving clock. Take the inertial proper time interval of the moving clock to be one unit so that you are plotting a relative time interval.*

3.1.1. Experimental Verification of the Metric Equation and Time Dilation

The somewhat bizarre concepts of length contraction and time dilation have been forced upon us by the assumption that the speed of light is constant in all inertial reference systems. Thus, if we accept Einstein's theory of special relativity, we would expect to see some evidence of length contraction and time dilation. This evidence can be found in the study of high-energy particles that are constantly raining down upon the earth's surface. When cosmic rays collide with atmospheric gases in the upper atmosphere muon's are produced which stream downwards toward the earth. These muons have a *laboratory* half-life of 1.52 μs (i.e., this is the half-life measured when the muons are at rest in the laboratory). Since some of the muons created in the upper atmosphere stream toward the earth with speeds of 0.99c, we should be able to test our ideas of time dilation.

Suppose we construct a muon detector which we carry to the top of a mountain. A similar detector is located at the foot of the mountain 1907 meters below. If the muon detector detects 100 counts per second at the top of the mountain, how many counts per second would we expect according to the classical picture? How many counts would our special theory of relativity predict? For muons traveling at a speed of 0.994c, the distance down the mountain would take 6.4 μs which is 4.2 half-lives. Since one-half of all the original muons decay in one half-life, we expect there to be $(0.5)^{4.2} = 0.054$ times the original muons, or approximately 5 muon counts per second detected at the bottom of the mountain. However, according to our theory of special relativity, the time interval measured in the frame of reference of the muon would be different from the time interval measured relative to the earth frame of reference. Using the metric equation, where $\Delta\tau_o$ is the lifetime of the muon measured in its own rest frame, and where Δt is the coordinate time interval in the earth frame and Δx is the distance interval measured in the earth frame, we have

$$\Delta\tau_o^2 = \Delta t^2 - \left(\frac{\Delta x}{c}\right)^2 \tag{24}$$

or

$$\Delta\tau_o^2 = (6.4\ \mu s)^2 - \left(\frac{1907\ m}{2.9979 \times 10^8\ m/s}\right)^2 = (0.704\ \mu s)^2 \tag{25}$$

This would imply that the muons have not even lived one full half-life as they move from the top of the mountain to the bottom [in fact, this is only 0.463 half-lives]. The relative number of muons surviving would, therefore, be $(0.5)^{0.463} = 72.5$, counts per second at the bottom of the mountain. Now 72.5 counts per second is quite a lot different from 5 counts per second, so that this experiment should be fairly definitive.

Experimental evidence supports the theory of special relativity. Far more muons are detected at ground level than would be predicted by non-relativistic arguments.

3.2. Relating *Distance* Measurements in Different Inertial Frames

As the light-clock moves through the laboratory frame, the initial light flash (sent out when the source is triggered) occurs at (t_A, x_A), and a second light flash occurs when the original light pulse is detected by the light-clock. We designate this event by the point (t_B, x_B). This means that the distance traveled by the light-clock as measured in the laboratory frame is given by $D = x_B - x_A$ so that the velocity of the light-clock through the laboratory frame is given by

$$v = \frac{x_B - x_A}{t_B - t_A} \tag{26}$$

Now the time interval measured between the emission of the light pulse and its subsequent detection in the reference frame attached to the light-clock is different from this same time interval measured in the laboratory frame, according to the equation

$$(t'_B - t'_A)^2 = (t_B - t_A)^2 - \left(\frac{x_B - x_A}{c}\right)^2 \tag{27}$$

so that the distance traveled by the light-clock as measured in the reference frame of the clock is given by

$$\begin{aligned}
D' &= v\,(t'_B - t'_A) = v\sqrt{(t_B - t_A)^2 - \left(\frac{x_B - x_A}{c}\right)^2} \\
D' &= \sqrt{v^2\,(t_B - t_A)^2 - \left(\frac{v}{c}\right)^2 (x_B - x_A)^2} \\
D' &= D\sqrt{1 - \left(\frac{v}{c}\right)^2}
\end{aligned}$$

This means that if there were a meter stick in the laboratory frame, and if the flashes of the clock/detector were observed at each end of that meter stick, an observer in the light-clock frame would claim that the distance between the ends of the meter stick was given by $D' < 1m$. Therefore, observers in two inertial reference frames traveling at a speed v relative to one another will disagree on the measurement of time intervals *and* length intervals.

EXERCISE 3.2 *Use MatLab to calculate and plot the measured length of a meter stick (i.e., a stick with a proper length of 1 meter) as a function of the relative speed β with which that meter stick passes through the laboratory.*

To summarize our conclusions concerning the special theory of relativity, we again examine the relationships between the measured time and length intervals in two inertial reference frames:

$$\Delta t' = \Delta t \sqrt{1 - \left(\frac{v}{c}\right)^2} \tag{28}$$

and

$$\Delta x' = \Delta x \sqrt{1 - \left(\frac{v}{c}\right)^2} \tag{29}$$

The first equation points out that time intervals, $\Delta t'$, as measured in a moving reference frame will always be *less than or equal to* the corresponding time interval measured in the laboratory frame. We define the time interval measured by a *single* clock moving at constant velocity as the *inertial proper time*, or the *spacetime interval*, and *we claim that this quantity will be the same for* ***any*** *inertial reference frame (i.e., it is a constant).* The time interval Δt, on the other hand, is a *coordinate time interval* which will depend upon the speed of the coordinate system and the separation of the two events as measured in that reference frame.

The second equation points out the distance between two points as measured in a frame of reference which is moving relative to those two points will always be *smaller* than the distance measured in a reference frame at rest relative to those two points. We define the distance between two points measured in the rest frame of the two points as the *proper length* or *proper distance* between the two points, and claim that this, too is a constant - having the same value for any inertial reference frame. This equation, therefore, states that the proper length of an object will always be *greater than or equal to* the length of the object as measured in any other inertial reference frame. The fact that objects observed in reference frames moving relative to an object will always measure the length of that object to be less than the length measured in the objects rest frame is known a Lorentz contraction.

The two equations taken together immediately show that the relative velocity measured in each reference frame is the same:

$$v' = \frac{\Delta x'}{\Delta t'} = \frac{\Delta x\sqrt{1-(v/c)^2}}{\Delta t\sqrt{1-(v/c)^2}} = v \tag{30}$$

EXERCISE 3.3 *Work problems 5 and 6 in Tipler.*

3.3. The Meaurement of Lengths *Perpendicular* to the Direction of Motion

Illustrate the necessity of perpendicular lengths being the same in all inertial frames [Moore pg. 65]

4. RELATING EVENT COORDINATES IN TWO INERTIAL REFERENCE FRAMES: THE LORENTZ TRANSFORMATION EQUATIONS

4.1. Derivation of the Lorentz Transformation Equations

We have shown that time and length intervals are different in different inertial reference frames, but that these differences are neglegible when the relative speeds of the reference frames are small. In this limit, the classical (Galilean) transformation is valid. It is only when the relative speeds of the reference frames become comparable to the speed of light that problems arise. To derive a general transformation equation for distance and time measurements we, therefore, begin with the equations that are approximately correct for low speeds (see Figure 1):

$$x = x' + vt \tag{31}$$

or

$$x' = x - vt \tag{32}$$

where we must determine if the quantity t in these equations is to be expressed as t or t'. If we treat these two equations as *transformation* equations, we would generally assume that the quantities on the right-hand-side should all be expressed in terms of the same coordinate system, giving

$$\begin{aligned} x &= x' + vt' \\ x' &= x - vt \end{aligned}$$

Since these last two equations are almost correct, but not quite, we *assume* that the correct transformations are of the form

$$x = \gamma(x' + vt') \tag{33}$$

and

$$x' = \gamma(x - vt) \tag{34}$$

Our task is to determine γ, which will necessarily be a function of the relative speed of the reference frames. We further assume that all coordinate clocks in the two reference frames are synchronized to $t = t' = 0$ when the origins of the two reference systems coincide and that a light flash occurs at the origin at this instant of time. Thus, because of the constancy of the speed of light in all inertial reference frames, we require that

$$x' = ct' \tag{35}$$

and

$$x = ct \tag{36}$$

Beginning with the equation for x, we have

$$\begin{aligned} x &= \gamma(x' + \frac{v}{c}ct') \\ &= \gamma(x' + \frac{v}{c}x') \\ &= \gamma x'(1 + \frac{v}{c}) \\ &= \gamma\left[\gamma(x - vt)\right](1 + \frac{v}{c}) \\ &= \gamma\left[\gamma(x - \frac{v}{c}ct)\right](1 + \frac{v}{c}) \\ &= \gamma^2 x(1 - \frac{v}{c})(1 + \frac{v}{c}) \end{aligned}$$

since $x = ct$. We can now solve for γ to obtain

$$\gamma = \frac{1}{\sqrt{1 - (v/c)^2}} \tag{37}$$

You will notice that $\gamma(v) \simeq 1$ for speeds much less than the speed of light, preserving Galilean relativity and the Galilean rule for the summation of velocities at these speeds. In fact, $\gamma(v)$ becomes significantly different from unity only for speeds greater than about 10% of the speed of light. This equation also seems to imply that it is not possible for the speed of the reference frame to *exceed* the speed of light, since γ would then become imaginary. We will examine this *cosmic speed limit* in more detail later.

Now that we know the form of γ, we need to find out how the time coordinate transforms. To do this we begin with

$$\begin{aligned} x' &= \gamma(x - vt) \\ &= \gamma\left[\gamma(x' + vt') - vt\right] \\ &= \gamma^2 x' + \gamma^2 vt' - \gamma vt \\ x'(1 - \gamma^2) &= \gamma^2 vt' - \gamma vt \end{aligned}$$

or, solving for the time t we obtain

$$\begin{aligned} t &= \frac{1}{\gamma v}\left[\gamma^2 vt' - x'(1 - \gamma^2)\right] \\ &= \gamma t' + \frac{\gamma^2 - 1}{\gamma v} x' \\ &= \gamma\left(t' + \frac{\gamma^2 - 1}{\gamma^2}\frac{x'}{v}\right) \\ &= \gamma\left[t' + \left(1 - \left(1 - \frac{v^2}{c^2}\right)\right)\frac{x'}{v}\right] \\ &= \gamma\left(t' + \frac{vx'}{c^2}\right) \end{aligned}$$

We have, therefore, derived the equations for transforming position and time coordinates from one frame to another moving with a relative speed of v:

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

You should satisfy yourself that the reverse transformation equations are given by

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

4.2. Measurement of Proper Length and Proper Time Intervals using the Lorentz Transformation Equations

We can use the Lorentz transformation equations to demonstrate the properties of time dilation and length contraction which we have mentioned previously. For example, we will consider the determination of the length of a meter stick in two inertial reference frames. In this particular example, we will assume that the meter stick is moving through the laboratory at a speed v. In the rest frame of the meter stick (which we will call the primed coordinate system) the length can be determined by measuring the positions of the ends of the meter stick at any arbitrary times we choose. We will let one end be located by x_1' and the other by x_2', such that the rest length (or proper length) of the meter stick is given by

$$L_o = x_2' - x_1' \tag{38}$$

Now the length of this same meter stick measured in the laboratory frame of reference can be determined precisely only by assuring that both ends of the meter stick are measured *at the same time*. Thus, only if the measurement of positions x_2 and $x_1$are made at the same instant in the laboratory frame of reference, can we assign a length to this meter stick, and that length, L, is given by

$$L = x_2 - x_1 \tag{39}$$

To examine this from the Lorentz equations, let's write our the length interval $x_2 - x_1$, using the Lorentz transformation equations. Here

$$x_2 = \gamma(x_2' + vt_2') \tag{40}$$

and

$$x_1 = \gamma(x_1' + vt_1') \tag{41}$$

so that

$$x_2 - x_1 = \gamma[(x_2' - x_1') + v(t_2' - t_1')] \tag{42}$$

But we know that the times at which we measure the location of the ends of the meter stick *in the moving reference frame* is totally irrelevant! So this equation is not too useful - what we needed was the one which contains t_2 and t_1! This latter equation gives

$$x_2' - x_1' = \gamma[(x_2 - x_1) - v(t_2 - t_1)] \tag{43}$$

where we must assure that $t_2 = t_1$, giving

$$x_2' - x_1' = \gamma(x_2 - x_1) \tag{44}$$

or

$$L_o = \gamma L \tag{45}$$

where L_o is the proper length measured in the rest frame of the meter stick, and L is the length of the meter stick measured in the laboratory frame of reference. This equation clearly shows that L is always less than or equal to L_o since $\gamma \geq 1$!

EXERCISE 4.1 *Problems 17, 18, and 19 in Tipler*

4.3. The Lorentz Velocity Transformation Equations and the Constancy of the Speed of Light

Since we now have a way of relating the measurement of time and distance intervals in two different reference frames, we can now determine how the velocity of an object in one reference frame is related to the velocity of that same object as seen in another reference frame moving with constant velocity relative to the first. We know from our earlier discussions that the velocity will not be simply additive as in the case of the Galilean velocity transformation, and that the correct transformation equations must give us the fact that the speed of light is the same in all inertial reference frames.

We begin by looking again at the Lorentz transformation equations. The x-component of the velocity of an object in any inertial reference frame must be the distance interval measured in that reference frame divided by the time interval measured in that same reference frame. We will use the symbol $\overrightarrow{u}$ to represent the velocity of an object as measured in an inertial reference frame, since we are using v to represent the velocity of one reference frame relative to another. Thus,

$$u_x = \frac{\Delta x}{\Delta t} = \frac{\gamma(\Delta x' + v\Delta t')}{\gamma(\Delta t' + \frac{v\Delta x'}{c^2})} \tag{46}$$

If we divide the top and bottom of the right-hand-side of this last equation by $\Delta t'$, we obtain

$$u_x = \frac{\Delta x}{\Delta t} = \frac{\gamma(\frac{\Delta x'}{\Delta t'} + v)}{\gamma(1 + \frac{v(\Delta x'/\Delta t')}{c^2})} = \frac{u'_x + v}{1 + u'_x v/c^2} \tag{47}$$

This equation is essentially equivalent to the Galilean transformation equation at low velocities, since the denominator is approximately equal to unity in these cases. However, for higher velocities, the denominator becomes significant.

Similarly, the velocity of an object moving in the y and z-directions are given by

$$u_y = \frac{\Delta y}{\Delta t} = \frac{\Delta y'}{\gamma(\Delta t' + \frac{v\Delta x'}{c^2})} = \frac{u'_y/\gamma}{1 + u'_x v/c^2} \tag{48}$$

and

$$u_z = \frac{\Delta z}{\Delta t} = \frac{\Delta z'}{\gamma(\Delta t' + \frac{v\Delta x'}{c^2})} = \frac{u'_z/\gamma}{1 + u'_x v/c^2} \tag{49}$$

We see that the velocity of an object measured in a direction perpendicular to the relative motion of the coordinate systems is dependent upon the velocity measured in the direction of motion of the coordinate system. The reason for this is that the time interval measured in one reference frame in not the same time interval measured in another. In the following exercises, you will see how to apply these velocity transformation equations and that they preserve the speed of light in all reference frames.

EXERCISE 4.2 *Problems 29 and 30 in Tipler.*

5. THE COSMIC SPEED LIMIT: A CONSEQUENCE OF THE CONCEPT OF CAUSE AND EVENT

5.1. Time Reversals and the Implication of Cause and Effect (Tipler Problem 82)

Consider two events A and B that occur at coordinate locations x_A and x_B and at times t_A and t_B, respectively in the laboratory or **Home** reference frame. Event A is the push of a button which initiates a light pulse, whereas event B is the reception of that light pulse which triggers the detonation of a bomb.

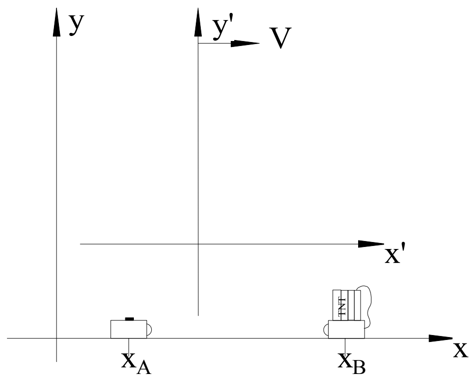

FIG. 10: Cause and effect can be demonstrated effectively by a light-activated explosion.

Clearly in the Home Frame, one event causes the other. If the laws of nature are to hold true in all inertial frames of reference, we *demand* that all inertial frames "see" the same events occuring *in the same order*, and not reversed. The reverse of cause and effect would have *profound* implications.

We now want to examine the Lorentz transformation equations to see what implications this requirement might make on our mathematical representation. We let the distance between the two events as measured in the laboratory frame be $D = x_B - x_A$ where $x_B > x_A$. We let the time interval between these same two events be $T = t_B - t_A$, where we *require* that $t_B > t_A$. If we calculate the time interval between these same two events as seen in a reference frame that is moving with speed v along the positive x-axis we obtain

$$\begin{aligned}
(t'_B - t'_A) &= \gamma\left[(t_B - t_A) - \frac{v}{c^2}\,(x_B - x_A)\right] \\
(t'_B - t'_A) &= \gamma\left[(t_B - t_A) - \frac{v}{c^2}\,c(t_B - t_A)\right] \\
(t'_B - t'_A) &= \gamma\,(t_B - t_A)\left[1 - \left(\frac{v}{c}\right)\right]
\end{aligned}$$

Now we know that $\gamma \geq 1$, so the time intervals $(t'_B - t'_A)$ and $(t_B - t_A)$ have the same sign, provided *the term in brackets is positive*, assuring that events occur in the primed frame in the same order that they occur in the unprimed frame. This means that

$$\left(\frac{v}{c}\right) \leq 1 \tag{50}$$

which implies that *the speed of any inertial reference frame which will preserve cause and effect must be less than or equal to the speed of light*! If any coordinate system could travel faster than c the whole framework of cause and effect would be unraveled.

So what about the case where two events A and B are separated by a distance Δx_{AB} which is *greater* than the distance that can be covered by light in the time interval Δt_{AB}? If no signal can be sent between these two events with a speed greater than the speed of light, then the two events *cannot be linked by cause and effect* - they are just two separate, independent events in space-time.

5.2. A Second Look at Space-Time Diagrams and World-Lines

We can summarize some of the concepts developed in the last section by examining a space-time diagram which illustrates how different events can be related to each other. We let one event A be located at the origin (at $x = 0$, $t = 0$). A second event B occurs some time t_B later at a position x_B, as seen in Figure 11. Events A and B *might* be two points along the path of a particular object that moves through space-time (this is a cause and effect relationship). You are familiar with position-time graphs in introductory physics. The slope of the curve on a position time graph at any instant in time is the instantaneous velocity of the object whose motion is represented by that curve. The same type thing is true for our space-time diagrams. Here, however, since the time and position axes have been swapped, the *inverse slope* of the line is the velocity of the particle (measured relative to the speed of light). To see how this works, consider the straight line connecting points A and B in the diagram. The slope of the line is given by

$$slope = \frac{c\,(t_B - t_A)}{(x_B - x_A)} = \frac{c}{\Delta x / \Delta t} = \frac{c}{u_x} \geq 1 \tag{51}$$

where u_x is the velocity of the particle. You can see that if u_x is less than c the slope of the line will be greater than unity.

Thus, objects moving through space-time at speeds less than the speed of light can be represented with "worldlines" which trace out the events along the particles path. These worldlines *cannot* have a slope anywhere which is *less* than unity, since that would imply that the object was moving *faster* than the speed of light. Since the two events A and B can be connected by a single line with slope greater than 1, these two events can indeed be two points on the worldline of the particle. Another way of saying this is that it is possible for information to pass from point A to point B so that what occurs at A may influence what occurs at B (i.e., A and B are related by cause and effect). Now since B occurs later in time than A, B occurs in A's future, whereas point C occurs in A's past. Now, consider a point on the positive x axis (which we will call point D, and which does not appear on the diagram above). This point *cannot* be connected with the origin with a line whose slope is greater than unity, thus, the two points A and D cannot be associated with cause and effect!

Thus, we divide the space-time diagram into regions: a cone-shaped region whose open end points in the $+t$ direction which we call Future, a cone-shaped region whose open end points in the $-t$ direction which we call Past, and a doughnut-shaped region whose open end points away from the origin, which we call Present. The Future and Past can be associated with the origin by cause and effect, but the Present cannot be!

EXERCISE 5.1 *In Diagram 12 three different worldlines are indicated which connect events A and B. Which of these are possible worldlines for a particle? Be sure to justify your answer.*

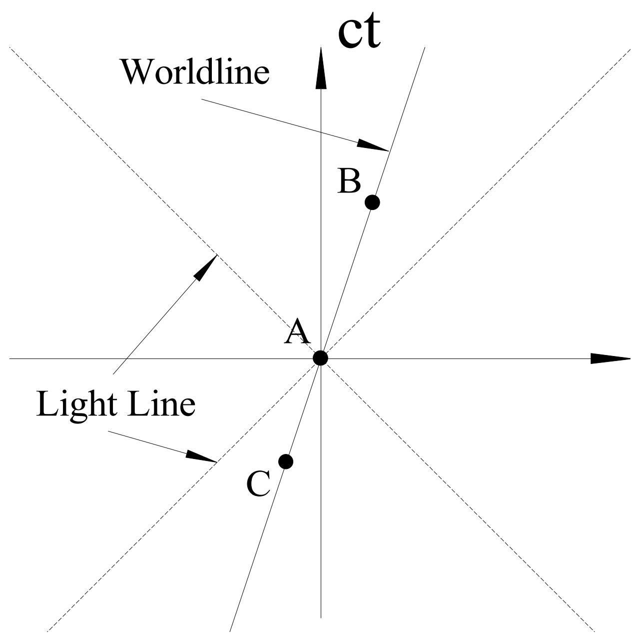

FIG. 11: Events which may be related to an event at the origin by cause and effect must lie within the light cone of that event—B being a *future* event, C a *past* event. Points which cannot be related to an event at the origin by cause and effect lie outside this light cone; not in the future or the past, but in the *present*.

5.3. Coordinate Transforms and Two Observer Space-Time Diagrams

These Lorentz transformation equations can be put in a more symmetric form by changing all time intervals to equivalent distance intervals (i.e., by multiplying the time by c). In this form we write

$$\begin{array}{ll} x = \gamma(x' + \beta ct') & x' = \gamma(x - \beta ct) \\ y = y' & y' = y \\ z = z' & z' = z \\ ct = \gamma(ct' + \beta x') & ct' = \gamma(ct - \beta x) \end{array} \tag{52}$$

where $\beta = v/c$.

We would like to see if we can map these transformations on a space-time diagram. Remember that a space-time diagram in a diagram of the spacial and temporal points of an event as seen in a given inertial reference frame. The spacial and/or temporal coordinates are typically scaled so that the light line has a slope of unity. This means that *distances* are represented in units of light-seconds, light-minutes, or light-years when time is represented in second, minutes, or years; or that distances are measured in meters while time is actually represented in a distance measurement (ct) called c-seconds, c-minutes, or c-years, with $c \simeq 3.0 \times 10^8\ m/s$. In fact, it is often convenient to define a set of units where $c = 1$, so that time and distance both have the same units.

The Lorentz transformation equations give us the relationships between times and positions measured in one inertial reference frame relative to the times and positions measured in another inertial frame. In particular, we can use the equation $x' = \gamma\,[x - \beta\,(ct)]$ to determine the location of all points on the space-time diagram for which $x' = 0$. The location of all such points is the ct' axis. Thus, the ct' axis is a line such that $x - \beta\,(ct) = 0$, or

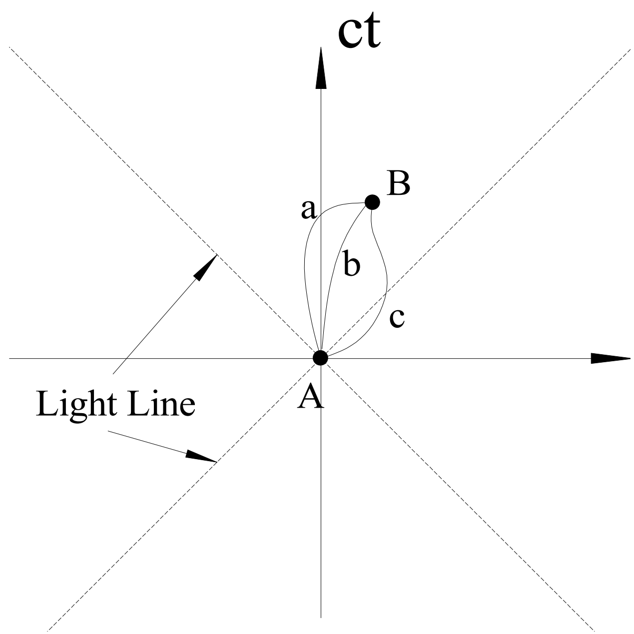

FIG. 12: Which of the three different "world lines" connecting events A and B can be realized in nature?

$ct = (1/\beta)\,x$, which is the equation of a straight line on the ct vs. x graph (see Fig. 13) with a slope of $1/\beta$ (i.e., where $\tan\theta = 1/\beta$). Likewise, we can use the equation $ct' = \gamma(ct - \beta x)$ to determine the location of all points on the space-time diagram for which $t' = 0$. The location of all such points is the x' axis. Thus, the x' axis is a line such that $ct - \beta x = 0$, or $ct = \beta x$, which is the equation of a straight line on the ct vs. x graph with a slope of β (i.e., where $\tan\phi = \beta$).

Plotting the $x' = 0$ and $ct' = 0$ lines on our space-time diagram result in what we call a two-observer space-time diagram. An event in space-time is represented by a single point on this graph. This point is the same for *both* reference frames, but the *coordinates* are not the same. You should recall that points which are equidistant from the origin along the x (or x') axis are always *parallel* to the time axes, and points which are equitime from the origin along the ct (or ct') axis are always *parallel* to the position axes. This means that transformations from one inertial frame to another are *not* orthogonal transformations as can be clearly seen in Figure 13. Rather, the ct' and x' axes are symmetrical about the light line, and both the x' and ct' axes move toward the light line with an increase in the relative speeds of the two inertial frames. The simplest example of a two-observer, space-time diagram is a plot of the location of the origin of a moving reference system through the laboratory. As you can see in Figure 14 , the origin traces out a line whose slope is inversely proportional to the speed of the reference frame. The origin moves in the positive x direction a distance Δx in the *time* $c\Delta t$. The location of the origin in the moving reference frame, however, never changes. It remains on the $x' = 0$ line for all time, but we can trace out the time in the moving reference frame along the $c\Delta t'$ axis.

Many of the interesting consequences of special relativity that we have discussed result from the non-orthogonality of the Lorentz transformations: 1) events simultaneous in one reference frame are not simultaneous in another, 2) moving clocks seem to run slower, and 3) the measured length of a moving object is always equal to or shorter than the proper length

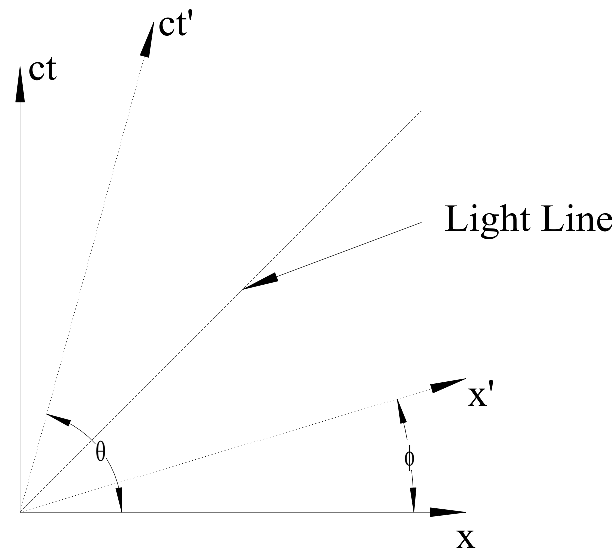

FIG. 13: A two-observer space-time diagram. A single event is plotted on this diagram as a single point. The coordinates of that single event, however, are measured to be different in two different inertial reference frames. The strange implications of relativity arise from the fact that these two coordinate systems are not orthogonal.

of the object. We want to see how these different phenomena can be understood based upon the two-observer space-time diagrams. The first of these aspects, that events simultaneous in one reference frame are not simultaneous in another, is relatively easy to see.

5.3.1. The Relativity of Simultaneity

As we have already pointed out, events which are simultaneous in one reference frame are *not* simultaneous in another. This can be clearly illustrated using the two-observer space-time diagram. In Figure 15, the two events R_1 and R_2 occur at the same time as observed in the **Home** frame (i.e., the ct vs. x frame). These two events, however, do *not* occur at the same time in the moving reference frame, where event R_2 (which is farther from the origin) appears to occur *before* event R_1. Likewise, the two events Q_1 and Q_2 which occur at the same time in the *moving* reference frame, do not occur at the same time as seen in the **Home** frame.

5.3.2. Calibration of the Two-Observer Space-Time Diagram

Length contraction and time dilation are a bit more difficult to see on these two-observer, space-time diagrams. To see these effects clearly, we must *calibrate* the axes. To accomplish this we need to find some quantity that does not change from one coordinate system to another. One such quantity is the space-time interval we introduced earlier. The space time interval is the time interval between two events which is measured by a single clock moving at constant velocity, i.e., an inertial proper time interval. The defining equation for the

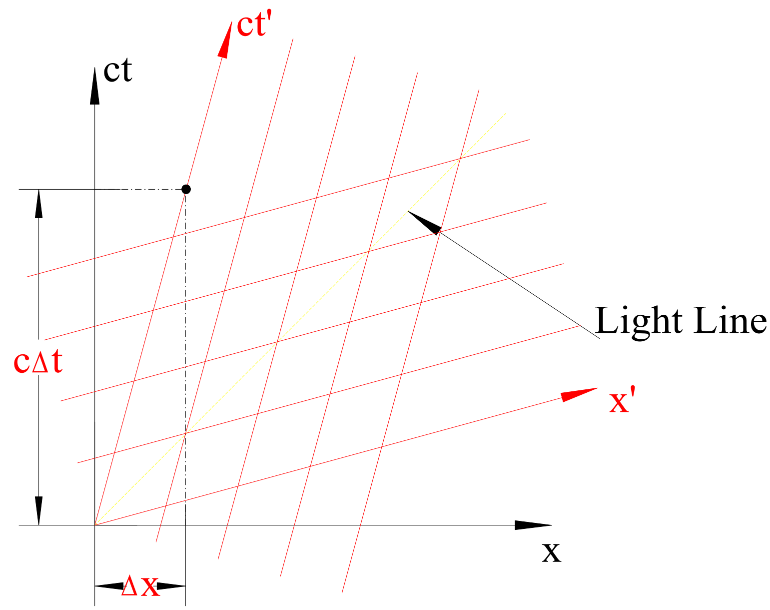

FIG. 14: The location of the origin of a moving reference frame as a function of time plotted in the two-observer space-time diagram. Notice that the origin moves a distance Δx in the rest frame and remain on the $x' = 0$ axis in the moving reference frame.

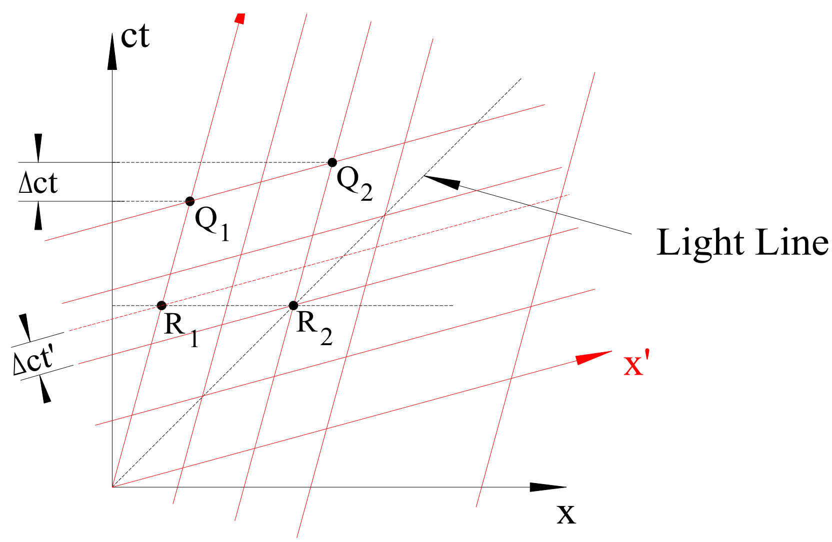

FIG. 15: Two events occurring one after the other in one inertial frame may occur in the reverse order in another inertial frame.

spacetime interval is

$$\Delta\tau_o^2 = \Delta t^2 - \frac{\Delta x^2 + \Delta y^2 + \Delta z^2}{c^2} = \Delta t^2 - \frac{\Delta R^2}{c^2} \tag{53}$$

and we can show that this space time interval has the same value in all inertial reference frames.

EXERCISE 5.2 *Show that the space-time interval is indeed an invariant by using the Lorentz transformation equations to show that* $(c\Delta t)^2 - \Delta R^2 = (c\Delta t')^2 - \Delta R'^2$.

We will assume that we want to calibrate the time and position scale relative to the origin where $x = x' = 0$ and where $t = t' = 0$, so that we can write the calibration equation in the form

$$(ct)^2 - x^2 = (ct')^2 - x'^2 \tag{54}$$

without the Δ's. To calibrate the ct' axis, we set $x' = 0$ and let $ct' = 1$. This gives us the equation for the hyperbola that pass through the ct axis at ± 1. Likewise, if we wish to calibrate the x' axis, we can set $ct' = 0$ and and let $x' = 1$. This gives us the equation for the hyperbola that pass through the x axis at ± 1, as shown in Figure 16. The points $P_1$and P_2 are the points on the x' and ct' axes that are one unit long.

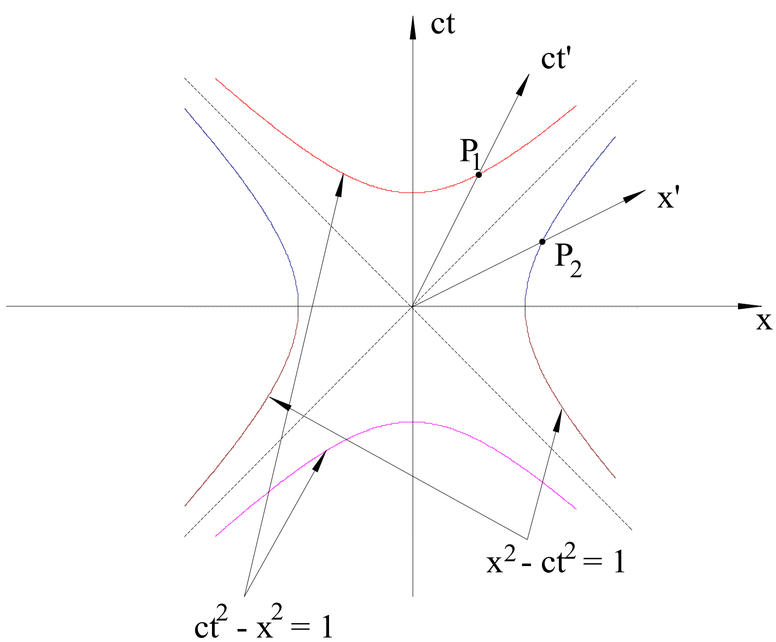

FIG. 16: Calibration of the two-observer space-time diagram using the invariant space-time interval. The scale is set equal to unity where the hyperbolic lines cross the position and time lines.

5.3.3. Length Contraction Illustrated with Two-Observer Space-Time Diagrams

We can illustrate the principle of length contraction by using this diagram and examining the "world lines" for an object which is one unit long (we will assume one meter) in the moving reference frame. These world lines will be parallel to the ct' axis since both ends of the "meter" stick are at rest in the moving reference frame. Since the hyperbolic calibration lines cross the x and x' axes at a point which must be the same length, we see in Figure 17 that the length of the meter stick is measured to be one meter in the moving frame, but is measured to be shorter in the **Home** frame, since both ends must be measured at the same time. For example consider the measurement of the length of the meter stick at time $t = 0$. Notice that as the velocity of the moving frame increases, the apparent length of the meter stick will get shorter!

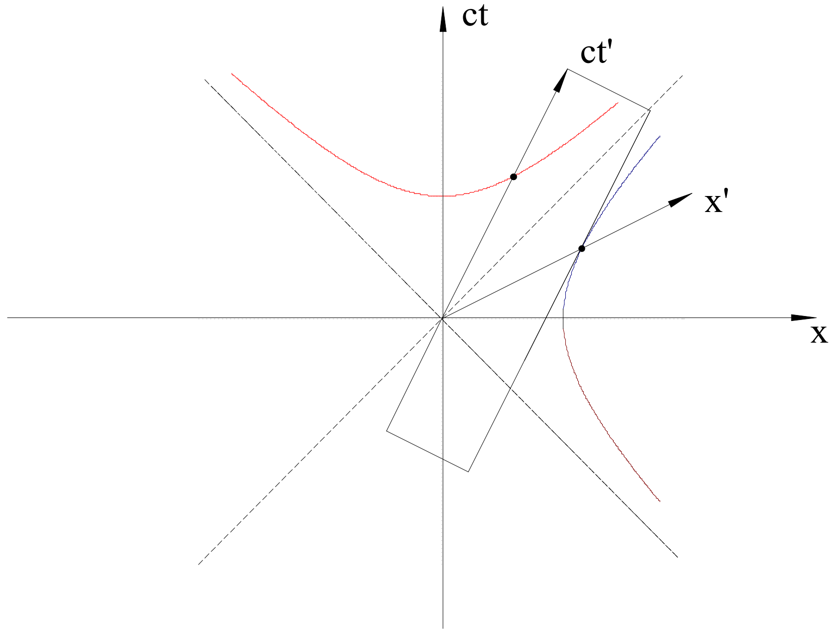

FIG. 17: Two world-lines are drawn for a "meter" stick which is at rest in the moving reference frame. One end is at $x' = 0$, while the other is at $x' = 1$. It is obvious that if both ends of this meter stick are measured at time $t = 0$ in the Home frame, the stick is less than one meter long!

EXERCISE 5.3 *Use a two-observer space-time diagram to illustrate the events which are discussed in the "Ladder and Barn" paradox. Let event A be the coinsidence of the front of the ladder with the front door to the barn. Let event B be the point where the front of the ladder is coincident with the back of the barn, and let event C be the point when the back of the ladder is coincident with the back of the barn. Show the locations and times as seen in the frame of reference of the barn and in the frame of the runner.*

5.3.4. Time Dilation Illustrated with Two-Observer Space-Time Diagrams

Similarly, we can illustrate the concept of time dilation using the two-observer space-time diagram. Consider the world-line of a clock which is moving through the Home frame of reference with a speed v. In the Home frame this clock is seen to change its location as the clock ticks. If the moving clock is set to zero as the clock moves past the origin of the Home frame, and if the clock at the origin of the Home frame is set to zero as the moving clock passed the origin, we have the situation which is depicted in Figure 18. Event A is where the moving clock passes the origin of the Home frame at time $t = 0$ as recorded in the Home frame. This same event occurs at time $t' = 0$ in the moving frame. Event B is when the moving clock has ticked off one unit of time (which may be a second, or a minute, or ...) Since the coordinates of the moving clock are not changing in the reference frame of this clock, the time line of the moving clock is just the ct' line drawn in our two-observer space-time diagram. You will notice that the time of event B as recorded in the Home frame (notice the dashed line parallel to the x-axis) is *later* than one unit of time. The time interval measured between events A and B in the Home frame is measured by two *different* clocks located at two *different* points in the Home frame. This is a coordinate time interval and

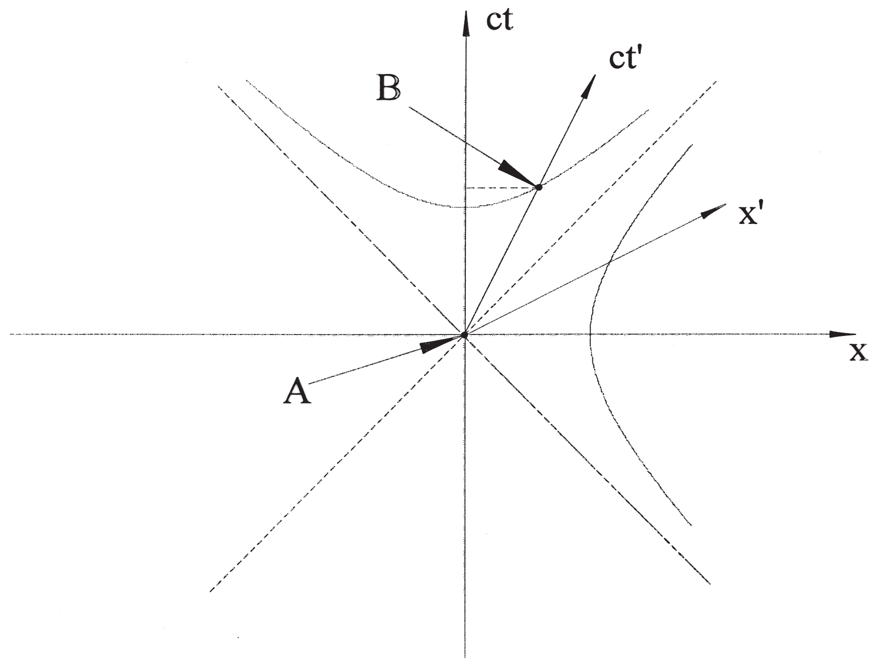

FIG. 18: Time Dilation - Proper Time in the Moving Frame.

is longer than the time interval as measured by the clock moving through the Home frame. This is consitent with our earlier discussion of time dilation: *moving clocks run slower*.

But what about a clock which is at rest in the Home frame? Will this clock appear to run slower in the moving frame? To see how this can be illustrated with the two-observer space-time diagram, consider Figure 19. Here, two events A and B are observed and recorded in the Home frame and in the moving frame. These two events occur at the same place (the origin) in the Home frame, but at different places as seen in the moving frame. The time interval as measured by a single clock at the origin of the Home frame is measured to be one unit long (one second, one minute, ...). This same time interval as measured in the moving frame is recorded by two different clocks at two different locations. This time interval is marked off by the dashed line in the figure which is parallel to the x' axis, and is clearly *longer* than one unit of time. So again we come up with the same result: time intervals between two events which are measured by a *single* clock are smaller than time intervals between these same two events measured by two different clocks.

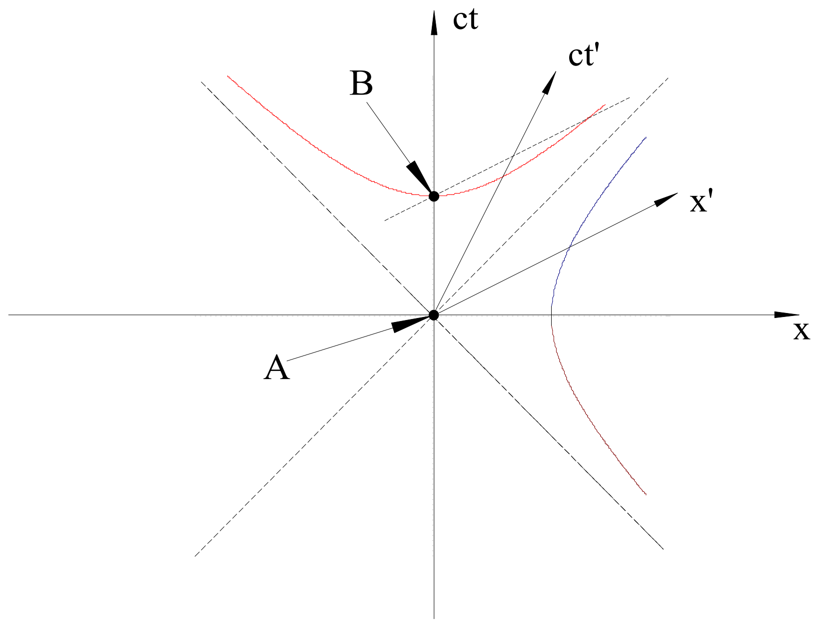

FIG. 19: Time Dilation - Proper Time in the Home Frame.

6. THE THREE TIME INTERVALS OF SPECIAL RELATIVITY

6.1. Time Intevals in Special Relativity Revisited

We have already discussed the fact that the measurement of the time interval between two events depends upon the reference frame in which the measurement is made. The metric equation

$$(c\Delta t)^2 - \Delta x^2 = (c\Delta t')^2 - \Delta x'^2 \tag{55}$$

makes this quite clear, since it shows that the time interval measured between two events in one frame depends upon the distance between the coordinates of these two events as measured in that frame. For this reason, we call this time interval a coordinate time interval. Since two events may be seen by many different observers moving in different inertial reference frames, each with a different speed, the distance between these two events will be measured to be different in the different reference frames. The coordinate time interval between these two events, then, must be different in each of the different frames of reference. Thus, there are a large number of coordinate time intervals which could be measured for the same two events.

A *special case* of the coordinate time interval occurs when two events occur *at the same location* in a particular inertial frame of reference. We define this time interval as the inertial proper time interval between the two events. We argued earlier that this time interval is an invarient, i.e., the measurement of this time interval is seen to be the same for all observers. As an example consider a motorist passing through an intersection. The motorist has a dashboard clock which is clearly visible to him *and to anyone who might look into his vehicle through the window*. In addition, there are two observers, each with a synchronized clock, standing at each end of this same intersection. As the motorist passes by the first observer (event A), that observer notes the time on his clock *and the time on the clock in* *the car.* As the motorist passes by the second observer (event B), he too notes the time on his clock and the time on the clock in the car. The coordinate time interval $t_B - t_A$ measured by the two observers at the intersection is just the difference between the times they noted on their individual clocks. The coordinate time interval $t'_B - t'_A$ measured by the motorist, however, is different because he is in a different reference frame. However, *the two observers will agree that the motorist clock read exactly what the motorist claims, because they could look through the window of the car and see just what the car clock indicated as it passed by*! Thus, *all* observers will agree on the time interval measured by the car clock, but the two observers who are standing at the intersection will measure a *different* time interval with their clocks. The time interval between the two events which is measured by the single clock is the invarient inertial proper time interval. Because it is an invariant, it is called the space-time interval, and is always the same *for all observers*.

Looking again at the metric equation, you will notice that the inertial proper time interval between any two events will always be *less* than the coordinate time interval between these same two events. So the inertial proper time interval is a special coordinate time interval, and in particular is the shortest possible coordinate time interval. This is represented in Figure 20.

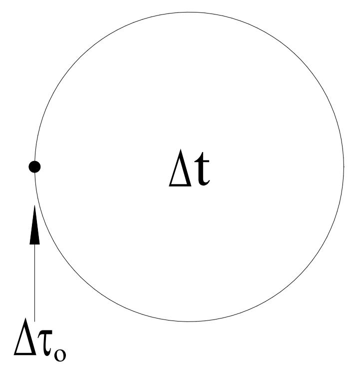

FIG. 20: There are as many different coordinate time intervals between two events as there are inertial reference frames; but there is only one inertial proper time interval—and it is the shortest time interval of all.

The fact that time intervals between two events may be measured to be different in different coordinate systems seems a bit strange at first, but there is a similarity with this situation and one with which we are a bit more familiar. The difficulty arises from the fact that distances and times are both inter-related in special relativity, and we don't normally connect these two. However, we are quite familiar with changes in spacial coordinate systems giving rise to changes in the components of vectors. Let's review these concepts for a moment and see how they may help us to understand the different time interval measurements encountered in special relativity.

6.2. A Geometric Analogy

Consider a city which is layed out with streets running primarily North and South, East and West. We can use an x, y coordinate system to unambiguously designate any point in the city, as shown in Figure 21 . Here we have defined the x-axis as the axis pointed toward

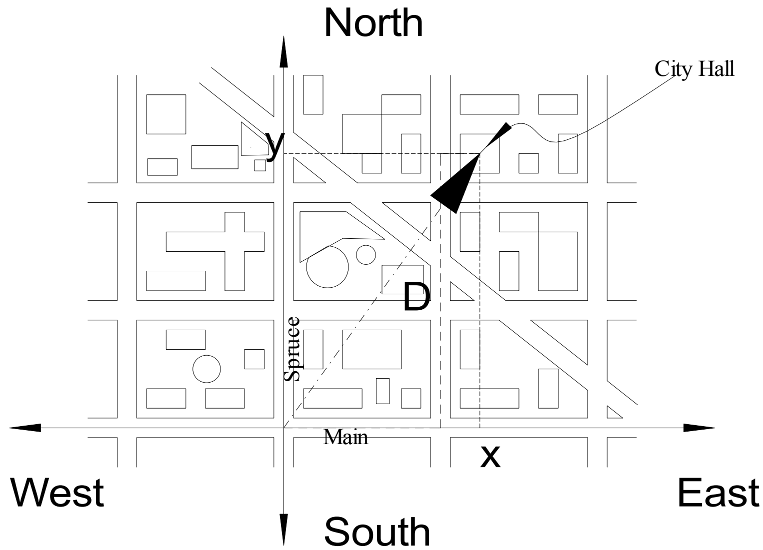

FIG. 21: The coordinates (x, y) and the displacement D of City Hall relative to the intersection of Main and Spruce are displayed on an East-West, North-South reference grid. One possible path from the intersection to City Hall is also designated by a dashed line.

the East and the y-axis as the axis pointed toward the North, with the origin arbtrarily chosen to be the center of Main and Spruce streets. The location of City Hall can thus be given by the point (x, y). If another coordinate system is chosen with the axes tilted at 45 degrees (see Figure 22), the location of City Hall will be represented by a new set of coordinates $(x', .y')$.

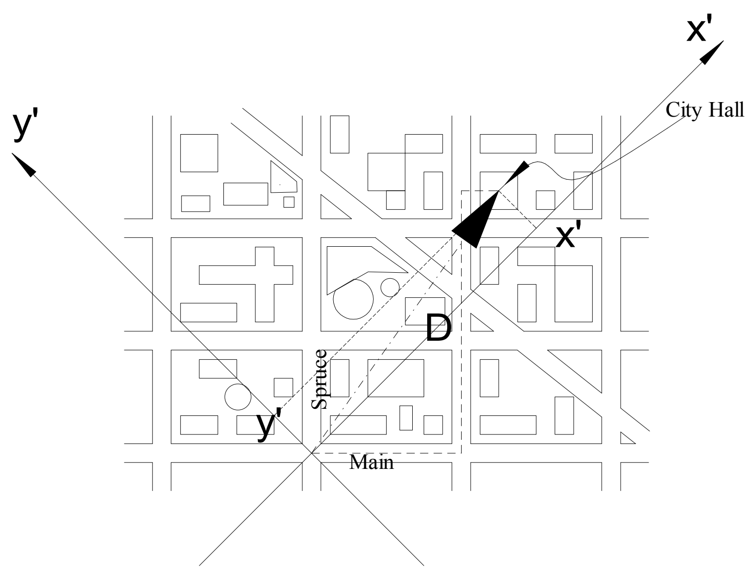

FIG. 22: The coordinates (x', y') and the displacement D of City Hall relative to the intersection of Main and Spruce are displayed on a reference grid oriented at 45 degrees from true North. One possible path from the intersection to City Hall is also designated by a dashed line.

However, the actual location of City Hall - its relationship to the other buildings in town - *does not change*, nor does the distance from City Hall to the intersection of Main and Spruce. In fact this distance can be defined in one of two ways: 1) the direct "as the crow flies" distance (which we might designate as a location or displacement vector D), or 2) the "path-length" distance traced out as you follow specific directions to reach City Hall along different streets (indicated by a dashed line in the figure). It should be obvious that the actual distance to City Hall, nor the path-length distance to City Hall change with a change in coordinate system. But it should also be obvious that the path-length distance will be different for different chosen paths. Each of the three concepts we use to specify the location of City Hall (coordinates, distance, and path-length distance) corresponds to a way of measuring time in relativity.

However, we have not yet seen a time interval in special relativity which corresponds to the concept of the path-length difference which depends upon the actual path taken to get to City Hall. Since this path-length difference should be the same no matter what coordinate system you choose, we might expect that there is some correspondence between the path-length difference and the proper time interval.

6.3. The Three Time Intervals of Special Relativity

Just as there are three ways to describe the location of City Hall geometrically, there are three different time intervals which must be distinguished in special relativity. These are:

1. The coordinate time interval measured in special relativity corresponds to the coordinates of City Hall - they change depending upon the particular coordinate system chosen. The *coordinate time* is the time which is read on a clock located at the place where the event occured. This clock must be synchronized to the clock at the origin of the coordinate system and thus defines a coordinate time interval relative to the origin clock. Coordinate times and time intervals between events must therefore depend upon the particular inertial reference frame chosen to represent the events.

2. The actual distance to City Hall, which is the same no matter what coordinate system we use, corresponds to the *inertial proper time interval*. It is always measured to be the same no matter who measures it. It is the time interval between two events as *measured by a single clock* which is moved through space-time *at constant speed* in such a way that it is at the same location as the two events when these events occur.

3. Like the path-length distance in our geometric analog, the *proper time interval* between two events may well depend upon the exact *world-line* which is chosen to connect these two events, as seen in Figure 23. (The clock does not *have* to move at constant velocity!) The *proper time interval*, just like the path-length distance, *has the same value for all inertial reference frames*, but will depend upon the exact path taken. Since the clock which measures the proper time interval is at the same location as the two events being measured, a person standing in the rest frame of the event will "see" the face of the moving clock at the instant of the event and will agree with a person who may be riding along in the coordinate system of the clock as to what that clock reads, just like the case with the car passing through the intersection. However, *the moving clock may not be moving at constant velocity in which case the time interval is not the inertial proper time interval, and is not an invariant*! And so we have a

proper time interval which is different from the inertial proper time interval. We need to consider how these two time intervals are related.

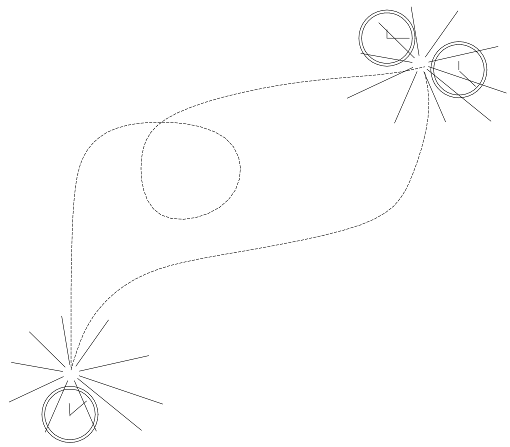

FIG. 23: The proper time interval measured by a single clock depends upon the world line traveled by that single clock. Two events may be measured by two different clocks which are located at the same two events but which travel between these two different events along very different world lines. Each clock will measure a proper time interval, but each will measure a different proper time interval!

6.4. Non-Inertial Proper Time Intervals

In Figure 24 , we have plotted the path of two space craft which move away from Earth (event A) and after some time return to Earth (event C). Each of the space craft travel to some distant star or space station along an inertial path. When they arrive at their destination, they immediately turn back toward Earth and return at the same speed, along yet another inertial path.

Let's first consider path ABC. This space craft travels to a point B which is two light years away from Earth and takes 4 years (as measured in the Earth's frame of reference) to reach there. The space craft, therefore, travels at a speed of $c/2$, as measured in the Earth's frame. He then returns to Earth at the same speed and arrives 8 years from the time he departed (as measured in the Earth's frame of reference). On board the space ship, the clock which rides along with the ship measures the proper time interval between these same two events, and obtains quite different results. For the trip out to point B we can determine the time interval on the ship's clock by using the invariant space-time interval

$$(c\Delta t)^2 - \Delta x^2 = (c\Delta t')^2 - \Delta x'^2 \tag{56}$$

In the reference frame of the ship, $\Delta x' = 0$, and the space-time interval $c\Delta t'$ (measured in an inertial reference frame) is given by

$$c\Delta t' = \sqrt{(c\Delta t)^2 - \Delta x^2} = \sqrt{4^2 - 2^2} = \sqrt{12} = 2\sqrt{3} \tag{57}$$

The same time interval occurs as the ship returns to Earth, so that the total time ellapsed on the ship's clock is $2(2\sqrt{3}) = 4\sqrt{3} = 6.93$ years. This means that the people on the ship aged by only 6.93 years while the people of the earth aged 8 years.

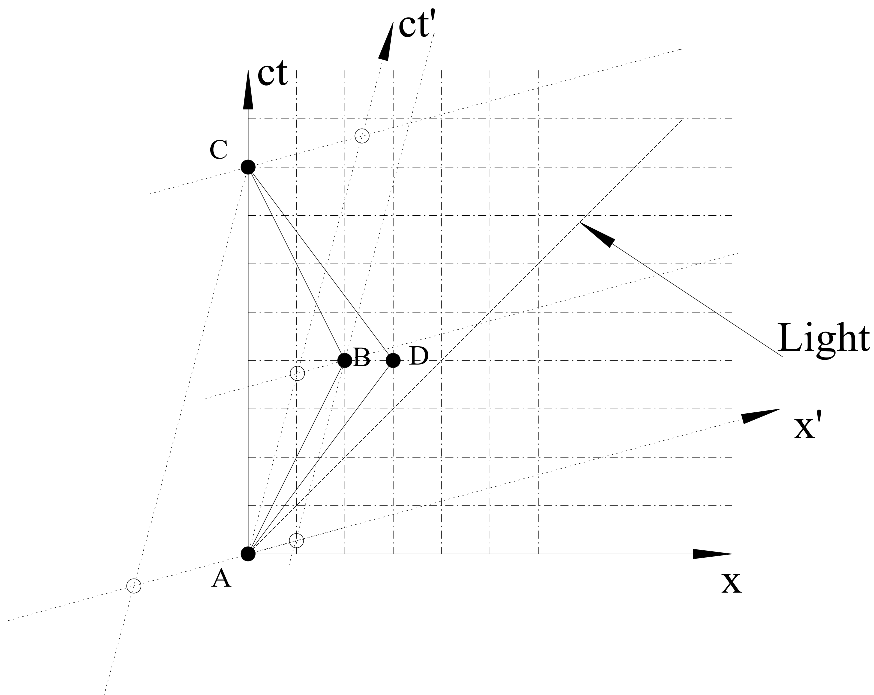

FIG. 24: Two non-inertial world lines taken by two different clocks. Each clock measures a "different" *proper time* interval.

Now consider path ADC for the second space craft. This space craft travels to a point D which is three light years away from Earth and takes 4 years (as measured in the Earth's frame of reference) to reach there. The space craft, therefore, travels at a speed of $3/4c$, as measured in the Earth's frame. He then returns to Earth at the same speed and arrives 8 years from the time he departed (as measured in the Earth's frame of reference). We again calculate the time interval measured by the on-board clock using the invariant space-time interval

$$(c\Delta t)^2 - \Delta x^2 = (c\Delta t')^2 - \Delta x'^2 \tag{58}$$

Again, $\Delta x' = 0$, and the space-time interval $c\Delta t'$ (measured in an inertial reference frame) is given by

$$c\Delta t' = \sqrt{(c\Delta t)^2 - \Delta x^2} = \sqrt{4^2 - 3^2} = \sqrt{7} \tag{59}$$

The same time interval occurs as the ship returns to Earth, so that the total time ellapsed on the ship's clock is $2(\sqrt{7}) = 5.29$ years. This means that the people on this ship aged by only 5.29 years while the people of the earth aged 8 years, and the people of the other ship aged 6.93 years.

The clocks on Earth which measure the time interval between the departure of the space craft and their return are inertial clocks, so that the inertial proper time interval between these two events is measured to be longer than the non-inertial proper times measured by the ship clocks. This means that we can have a number of different possible proper time intervals between two events, each of which depends upon the worldline of the individual clocks (path ABC or ADC). But each non-inertial proper time interval is *shorter* than the inertial proper time interval (the space-time interval). Thus, the relationships between the different time intervals within special relativity that we have discussed can be summarized by the following diagram (Figure 25).

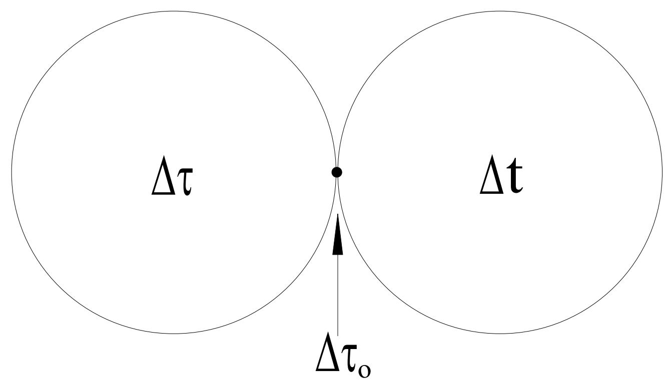

FIG. 25: A schematic showing how the three different time intervals which arise in special relativity are related.

Notice that in the second case above, the space craft moves a distance of 6 light years (three out and three back) in a time of 5.29 years, or 1.13 c! This would seem to violate our assumption that nothing can move faster than the speed of light. However, we have not been careful in the way we use our numbers. We have used the distance traveled as measured in the Earth frame of reference and the time as measured in the ship frame of reference. This is an illegal operation. If you use the Lorentz transformation equations to find the distance traveled in the ship frame (or the time traveled in the earth frame) you will get the correct speed for the spacecraft, $3/4c$ in this case, and it will always be less than c.

Now, we can do this same thing again with another space ship moving even faster, say with a speed of $0.996c$. Here, the time interval as measured on the ships clock is *much much less* than the time interval measured in the Earth's frame of reference. This will mean that the "apparent speed" of the ship will be much much greater than the speed of light. If we carry this scenario further, you will realize that it is possible to go to any point in the universe within a finite length of time (as measured on the ship's clock) provided you can travel at speeds very near the speed of light. This means that if the ship were to move at the speed of light, then the time interval as measured by the ship's clocks would become zero, so that a resident in the space ship would not age any as he moved across the galaxy! Remember that light travels at the speed of light, so that photons must be able to traverse the entire galaxy in no time at all!

6.5. The Twin Paradox

The twin paradox illustrates the differences between proper time and *inertial* proper time. The person traveling in the spaceship measures proper time, but the person who remains on Earth measures the inertial proper time. The inertial proper time is always greater than or equal to a non-inertial proper time. [For a full discussion, see Moore.]

7. MOMENTUM AND ENERGY CONSERVATION IN SPECIAL RELATIVITY.

7.1. The Failure of Newtonian Momentum Conservation

One of the most fundamental conservation laws in classical physics is the law of conservation of momentum. If the special theory of relativity is valid then the conservations laws must be valid in all inertial reference frames. Since the Galilean velocity addition rule is *not* valid in the special theory of relativity, we might suspect there to be a problem with momentum conservation. The following exercise demonstrates that the classical Newtonian momentum is *not* conserved in all inertial reference frames.

EXERCISE 7.1 *Consider a particle of mass* m *moving in the* $+x$ *direction with a speed of* $v_{1i}/c = +3/4$ *in the Home Frame. This particlce collides with a particle of mass* $2m$ *which is at rest,* $v_{2i} = 0$*. The lighter particle rebounds from the collision with an* x *velocity of* $v_{1f}/c = -1/4$.

1. *Determine the velocity of the heavier particle assuming the conservation of Newtonian momentum for this system, and show that it rebounds with an* x *velocity of* $v_{2f}/c = +1/2$.
2. *Now consider another reference frame, the Other Frame, which is moving along the positive* x *axis of the Home Frame with a constant velocity,* $v/c = +3/4$*. Express the momentum of each of the particles in the Other Frame both before and after the collision, using the Lorentz transformation equation for velocities. Is the Newtonian momentum of the system conserved in this reference frame?*

The fact that the Newtonian momentum is not conserved in all inertial frames means that we must look for a *new* definition for the momentum which is consistent with the Newtonian momentum at low speeds.

7.2. A New Definition of Momentum

This new, more general, definition of momentum, therefore, must be something slightly different from the mass of an object times the object's velocity as measured in a given reference frame, but must be *similar* to the Newtonian momentum since we must preserve Newtonian momentum at low speeds. We will assume that the mass of an object is invariant (i.e., constant) in all inertial reference frames. This means that we need to redefine the velocity part of the momentum equation in such a way that the newly defined momentum is conserved in all reference frames, when we apply the Lorentz transformation.

One clue which might help us find the correct form for the momentum is the fact that the Lorentz transformation equations for position and time are *different* from the Lorentz transformation for velocities. The primary difference is that there is *no* change in the *perpendicular* components of the *position vector* from one inertial frame to another, while there *is* a change in the perpendicular components of the *velocity.* This difference is primarily due to the fact that *time intervals measured in one reference frame are not equal to time intervals measured in another frame of reference.* Thus, when we divide the change in y and z by the time interval, we introduce a term which makes the velocity equations not transform like the position coordinates.

If the momentum is to transform like the position, and not like velocity, we must divide the perpendicular components of the vector position by a quantity that is *invariant.* The logical quantity to try is the *space-time interval.* By dividing the components of a position vector by the space-time interval rather than the coordinate time interval, we obtain a quantity with the units of velocity whose perpendicular components transform like position components, but not like velocity components. The space-time interval can be written in the form

$$c\Delta\tau = \sqrt{(c\Delta t)^2 - \Delta R^2} \tag{60}$$

$$= c\Delta t\sqrt{1 - \left(\frac{\Delta R}{c\Delta t}\right)^2} = c\Delta t\sqrt{1-\beta^2} \tag{61}$$

where β is the velocity of the object moving through a given coordinate system measured relative to the speed of light. Since the space-time interval is invariant, we can also write

$$c\Delta\tau = \sqrt{(c\Delta t')^2 - \Delta R'^2} \tag{62}$$

$$= c\Delta t\sqrt{1 - \left(\frac{\Delta R'}{c\Delta t'}\right)^2} = c\Delta t'\sqrt{1-\beta'^2} \tag{63}$$

Now, if the displacement of an object measured in a given intertial frame is divided by the space-time interval, we obtain

$$\frac{\Delta x}{c\Delta\tau} = \frac{\Delta x}{c\Delta t\sqrt{1-\beta^2}} = \frac{u_x/c}{\sqrt{1-\beta^2}} \tag{64}$$

or

$$\frac{\Delta x}{\Delta\tau} = \frac{u_x}{\sqrt{1-\beta^2}} \tag{65}$$

We see that the quantity $\Delta x/\Delta\tau$ is essentially equivalent to the velocity of a particle measured in a given frame of reference if the velocity is small (i.e., if β is small). Thus, since the classical definition of momentum $m\,u_x$, is essentially the same thing as $m\Delta x/\Delta\tau$ for small speeds, we postulate that the correct form of the relativistic momentum is given by

$$m\frac{\Delta x}{\Delta\tau} = \frac{mu_x}{\sqrt{1-\beta^2}} \tag{66}$$

Notice that $1/\sqrt{1-\beta^2}$ looks like our definition of γ. However, in this equation β is *not* a measure of the speed of one inertial reference frame relative to another, but a measure of the velocity of a particle *measured in a single reference frame*. We must, therefore, distinguish between these two different γ's. We will write $\gamma(v)$ to represent the γ we have used in the Lorentz transformation equations and which depends upon the relative velocity between two reference frames. We will write $\gamma(u)$ to represent the γ associated with the velocity of a particle measured in a given reference frame.

Thus, we postulate that the correct expression for the momentum of an object of mass m is given by

$$m\frac{\Delta x}{\Delta\tau} = \frac{mu_x}{\sqrt{1-\beta^2}} = \gamma(u)mu_x \tag{67}$$

The fact that the space-time interval between two events is measured to have the same value in any two reference frames means that we can also write

$$m\frac{\Delta x'}{\Delta\tau} = \frac{mu'_x}{\sqrt{1-\beta'^2}} = \gamma(u')mu'_x \tag{68}$$

Similarly, we can write

$$\begin{aligned} m\frac{\Delta y}{\Delta\tau} &= \frac{mu_y}{\sqrt{1-\beta^2}} = \gamma(u)mu_y \\ m\frac{\Delta z}{\Delta\tau} &= \frac{mu_z}{\sqrt{1-\beta^2}} = \gamma(u)mu_z \end{aligned}$$

It is interesting to note that the ratio of the time interval in a given reference frame to the invariant time interval $\Delta\tau$ is given by:

$$\begin{aligned} \frac{\Delta t}{\Delta\tau} &= \frac{1}{\sqrt{1-\beta^2}} = \gamma(u) \\ \frac{\Delta t'}{\Delta\tau} &= \frac{1}{\sqrt{1-\beta'^2}} = \gamma(u') \end{aligned}$$

We now need to look at how this proposed definition of momentum will transform, to see if this will preserve our conservation rules. Beginning with the Lorentz transformation equations in terms of intervals

$$\begin{aligned} \Delta x &= \gamma(v)\left[\Delta x' + \beta(c\Delta t')\right] \\ \Delta y &= \Delta y' \\ \Delta z &= \Delta z' \\ c\Delta t &= \gamma(v)\left[(c\Delta t') + \beta\Delta x'\right] \end{aligned}$$

we divide both sides by the invariant space-time interval to obtain

$$\begin{aligned} \gamma(u)u_x &= \gamma(v)\left[\gamma(u)u'_x + \beta\gamma(u')c\right] \\ \gamma(u)u_y &= \gamma(u')u'_y \\ \gamma(u)u_z &= \gamma(u')u'_z \\ \gamma(u)c &= \gamma(v)\left[\gamma(u')c + \beta\gamma(u')u'_x\right] \end{aligned}$$

Multiplying through by m in each of these equations and defining the components of the momentum vector $\overrightarrow{p}$ to be

$$\begin{aligned} p_x &= \gamma(u) m u_x \\ p_y &= \gamma(u) m u_y \\ p_z &= \gamma(u) m u_z \end{aligned}$$

we obtain

$$\begin{aligned} p_x &= \gamma(v)\left[p'_x + \beta\gamma(u') mc\right] \\ p_y &= p'_y \\ p_z &= p'_z \\ \gamma(u) mc &= \gamma(v)\left[\gamma(u') mc + \beta p'_x\right] \end{aligned}$$

In addition to the three components of the momentum vector, the application of the Lorentz transformation equation has introduced a fourth quantity, $\gamma(u)mc$, which we need to identify. To do this we look at the classical limit of these four terms. As we have already mentioned, when the particle velocity is small, u/c is small and $\gamma(u) \to 1$, so that the relativistic definition of the momentum is equivalent to the classical definition. Likewise, in the limit as $\gamma(u) \to 1$ (i.e., when $\beta = u/c \to 0$) we see that the term $\gamma(u)mc$ becomes

$$mc\left[1-\beta^2\right]^{-1/2} \simeq mc\left[1+\frac{1}{2}\beta^2+\ldots\right] \simeq mc+\frac{1}{2}mu^2/c+\ldots \tag{69}$$

The second term in this expansion is *similar in form to the kinetic energy* of the particle! To make the similarity complete, we multiply this last equation by c and *define* the relativistic energy E of a particle to be

$$E = \gamma(u) mc^2 \tag{70}$$

Thus, the relativistic energy, in the limit as $\gamma(u) \to 1$, becomes

$$E = \gamma(u)mc^2 \simeq mc^2\left[1+\frac{1}{2}\beta^2+\ldots\right] \simeq mc^2+\frac{1}{2}mu^2+\ldots \tag{71}$$

For this last expression to be valid in the classical limit, a particle must have energy even when it is at rest (a *rest mass* energy). If mass in truely invariant, this constant rest mass would simply introduce a constant offset in the energy scale. This offset creates no problem since we are usually interested only in *changes* in the total energy of a system. The kinetic energy of a particle, then, in special relativity can be written as

$$K = \frac{1}{2}mu^2 = E - mc^2 = \gamma(u)mc^2 - mc^2 = \left[\gamma(u)-1\right]mc^2 \tag{72}$$

A more troublesome question arises if we take $\gamma(u)mc^2$ as an expression for the *total* energy of the system, since there is no potential energy term. In classical physics we normally write the total mechanical energy as the sum of the kinetic and the potential energy. If this is truely an expression for the total energy of the system, the potential energy must somehow be tied up with the inertial mass of the system in a way which has not been obvious in classical systems.

Let's make a wild assumption that a change in potential energy of a system is equal to the change in the mass of the system. If this were actually correct any change in mass of a classical system would have to be so small as to be immeasurable, even when a measurable change in the potential energy occured. The following example illustrates this change in mass for a typical situation.

Example 7.1 *Consider a 1 kg mass which is lifted through a distance of 1 meter. The gravitational potential energy of this mass has changed by an amount equal to 9.8 Joules. By how much would the mass of this system change according to our definition of the total relativistic energy? Since the system is at rest, any change must be in the rest mass energy, so that*

$$\Delta E = \Delta mc^2 \tag{73}$$

or, for the mass change

$$\Delta m = \Delta E/c^2 = \left(9.8\ /9.0 \times 10^{16}\right)\ kg \tag{74}$$

which means that we would have to measure mass changes to better than one part in 10^{16} *to see this effect!*

Thus, typical potential energy changes in classical physics would lead to negligible changes in the masses of the components of the system - they simply cannot be detected. There are cases, however, where this mass difference has been measured. This occurs when nuclear components (e.g., a proton and a neutron) bind together to form a more complex structure. The mass of the individual components is always *larger* than the mass of the two components in the more complex structure, implying that the mass change is what supplies the binding energy of the more complex structure.

EXERCISE 7.2 (Tipler, Problem 1.43) *A free neutron decays into a proton plus an electron:*

$$n \rightarrow p + e \tag{75}$$

Use Table 1-1 in Tipler to calculate the energy released in this reaction.

Having justified our definition of the energy as

$$E = \gamma\left(u\right) mc^2 \tag{76}$$

we can now write the momentum transformation equations in the form:

$$\begin{aligned} p_x &= \gamma(v)\left[p'_x + \beta\frac{E'}{c}\right] \\ p_y &= p'_y \\ p_z &= p'_z \\ \frac{E}{c} &= \gamma(v)\left[\frac{E'}{c} + \beta p'_x\right] \end{aligned}$$

These equations, taken together, are similar in form to the position-time four-vector in relativity. This new momentum-energy four-vector is sometimes called the momenergy. The fact that the momentum and energy are inter-related in special relativity would seem to indicate that principles of the conservation of momentum and of energy in Newtonian physics are actually more fundamental than we might have originally thought.

The following exercises illustrate the use of the conservation of momentum and energy in relativistic problems.

EXERCISE 7.3 *Two equal masses $m_1 = m_2 = m_0$ approach each other in the Home reference frame with equal speeds $u_o = 0.6c$ (choose m_1 to be moving in the $+x$ direction). These two masses collide and form a single mass M which is at rest in the Home frame.*

1. *Use the relativistic expressions for the momentum and energy in the Home frame to show that the momentum of the system is conserved and that conservation of energy requires that the final mass M must be greater than $2m_0$.*

2. *Consider another reference frame S which is attached to m_1. Show that the momentum and energy of the system are also conserved in this new reference frame and that the final combined mass M is the same as we obtained in the Home frame.*

EXERCISE 7.4 *A body of rest mass m_o, traveling initially at the speed of 0.6c, makes a completely inelastic collision with an identical body that is initally at rest.*

1. *What is the rest mass of the resulting single body?*

2. *What is its speed?*

EXERCISE 7.5 *(Extra Credit) Consider a particle of mass m moving in the $+x$ direction with a speed of $v_{1i}/c = +3/4$ in the Home Frame. This particlce collides with a particle of mass $2m$ which is at rest, $v_{2i} = 0$.*

1. *Using the equations we have developed for the relativistic momentum and energy, determine the velocity of the heavier particle after the collision. Is energy conserved in this collision?*

2. *Now consider another reference frame, the Other Frame, which is moving along the positive x axis of the Home Frame with a constant velocity, $v/c = +3/4$. Express the momentum and energy of each of the particles in the Other Frame both before and after the collision, and show that the relativistic momentum is conserved in this collision in the Other Frame. Is energy conserved in this reference frame?*

We have shown that the relativistic momentum is given by

$$\overrightarrow{p} = \gamma(u)\, m\overrightarrow{u} = \frac{m\overrightarrow{u}}{\sqrt{1-(u/c)^2}} \tag{77}$$

and that the relativistic total energy is given by

$$E = \gamma(u)\, mc^2 = \frac{mc^2}{\sqrt{1-(u/c)^2}} \tag{78}$$

Note that we can multiply the equation for $\overrightarrow{p}$ by c^2 and obtain the relationship

$$\overrightarrow{p}c^2 = E\overrightarrow{u} \tag{79}$$

or

$$\frac{\overrightarrow{u}}{c} = \frac{\overrightarrow{p}c}{E} \tag{80}$$

The fact that momentum and energy can be combined together as a four-vector, just like time and space should make you wonder if there is an invariant quantity which arises from this four-vector just as the invariant space-time interval arises from the space-time four-vector. In fact, there is!

Just as in the case of the space-time four-vector, we take the square of the fourth vector and subtract the sum of the squares of the other three vectors to obtain a new invariant *interval*. (Note: This can also be thought of as multiplying the fourth component of the momenergy by $i = \sqrt{-1}$, and taking the sum of the squares of all the terms.) This invariant interval has the same value in all inertial reference frames and is equal to $(mc)^2$, as shown below.

$$\begin{aligned}
(E/c)^2 - \left[p_x^2 + p_y^2 + p_z^2\right] &= \left[\gamma(u)mc\right]^2 - \left[(\gamma(u)mu_x)^2 + (\gamma(u)mu_y)^2 + (\gamma(u)mu_z)^2\right] \\
&= (\gamma(u)mc)^2 - (\gamma(u)m)^2\left[u_x^2 + u_y^2 + u_z^2\right] \\
&= (\gamma(u)mc)^2\left[1 - (u/c)^2\right] \\
&= (mc)^2
\end{aligned}$$

Thus, we have the relation (valid in all inertial reference frames)

$$(E/c)^2 - p^2 = m^2c^2 \tag{81}$$

which can be expressed as

$$E^2 = p^2c^2 + m^2c^4 \tag{82}$$

This equation has the form of the Phythagorean theorem applied to a right triangle, where the total energy E is the hypotinuse, the rest mass energy mc^2 is the base, and the term p^2c^2 is the vertical side of the right triangle. Thus, the vertical side corresponds roughly to the kinetic energy, depending upon the momentum of the particle.

This last equation is extremely significant, since it indicates that an object with no rest mass can still have a momentum! Thus, for a photon with zero rest mass, the momentum can be written as

$$p = E/c \tag{83}$$

We will show later that photons have energy $E = h\nu$, so that this last expression can be written

$$p = h\nu/c = h/\lambda \tag{84}$$

so that the momentum of a photon is directly related to the wavelength of the photon.