UCD: Physics 9HB – Special Relativity and Thermal/Statistical Physics

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

This text was compiled on 04/15/2025

TABLE OF CONTENTS

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

CHAPTER OVERVIEW

1: Foundations of Relativity

1.1: The Relativity Principle

Inertial Frames

Back in our studies of classical mechanics, we spent a very brief period of time learning about how to relate the measurements of position and time between two observers in relative motion (go here for the LibreText reminder of this topic). Actually "relative motion" in this context is imprecise. We restricted our study to a very specific kind of relative motion – that for which the two observers maintain a constant relative velocity.

For what is to come, we will restrict these frames of reference even further – we will insist that every observer makes its measurements from an *inertial frame*. This kind of frame is one in which Newton's first law assures that objects will not spontaneously begin to accelerate. That is, if we are in such a frame, and we eliminate all of the real forces present on a stationary object, then the object remains at rest in that frame.

The simplest way to understand inertial frames is to consider what kind of frame is *not* inertial. Suppose you are in spaceship (far away from all gravitational sources), and it is accelerating forward. You hold a pencil in your hand, which is at rest in your frame, but you are exerting a force on it with your fingers, so to test to see if you are in an inertial frame, you release it. As soon as you do, it continues with whatever speed it had at the moment of release, while you and your spaceship continue to accelerate. From your perspective, it is the pencil that accelerates, which tells you that you are not in an inertial frame.

[*One might wonder why an inertial frame is an additional restriction beyond what we did in 9A. Certainly the intention was to deal with inertial frames back then, but technically two frames with equal acceleration vectors (but unequal velocities) will also satisfy the Galilean transformation.*]

Postulate(s)

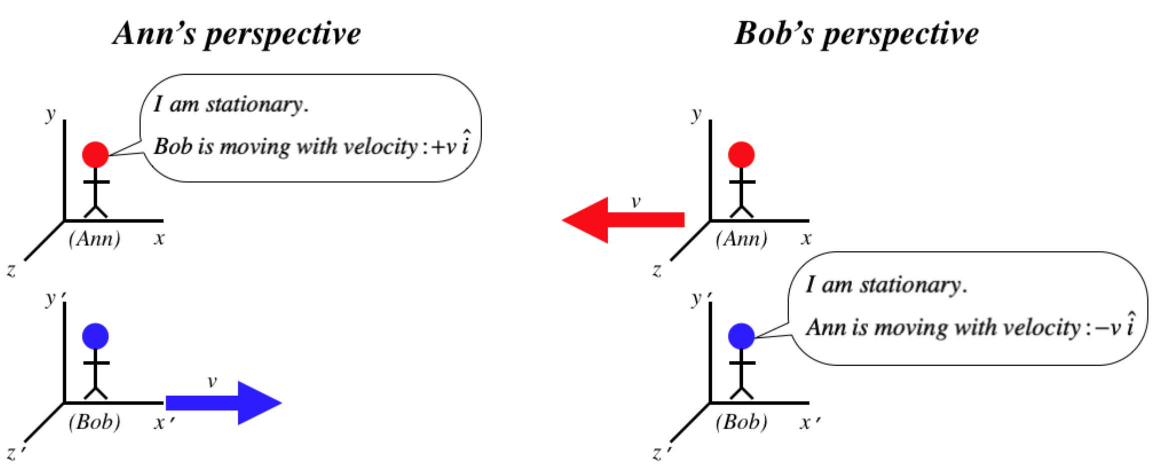
We have a simple experiment for testing whether our frame is inertial, but it doesn't tell us whether our frame is stationary or moving in a straight line at a constant speed, because when we release the pencil under these circumstances, it remains stationary from our perspective in both cases. So what kind of experiment will tell us whether or not we are moving?

Albert Einstein pondered this very thought, and came up with no answer. Eventually, he felt compelled to assert it as a fundamental aspect of our universe, and the *relativity principle* was born:

No experiment can be performed within an inertial frame that determines whether it is moving or at rest.

This is also known as the first postulate of the theory of Special Relativity. One way that we can express this is in terms of an "argument" between two observers.

Figure 1.1.1 – All Observers in Inertial Frames Can Claim to Be Stationary



[*These kinds of diagrams, where the perspectives of two observers in relative motion, will have some common elements. First, we will always define their relative motion to be parallel to their common x-axes. Second, we will define the primed frame to be moving in the $+x$ direction relative to the unprimed frame.*]

Calling this the "first postulate" implies that there is a second postulate, and there is, though one could argue that it follows directly from the first postulate and therefore doesn't need to be stated separately. It is this:

Every observer measures the velocity light to be the same value.

The reason this "second postulate" can be considered a consequence of the first postulate is that the theory from which we derive the speed of light contains no provisions for the motion of the observer (or rather, it predicts the same speed for all observers). Therefore the theory predicts that any experiment that measures the speed of light in a vacuum will give a specific answer. If different inertial frames produced different values for the speed of light, we would have a violation of the first postulate, as we would then have an experiment to determine the "true" rest frame. So to the extent that we accept this theory of light propagation, we don't need the second postulate.

Digression: "Ultimate Speed"

The discussion above actually paints a somewhat in inaccurate picture of the foundation of relativity. As we will see later, these postulates lead to the requirement of the speed of light being the limit which no relative motion can ever exceed. It turns out that if we just postulate that such an "ultimate speed" exists, then relativity results, independent of the theory of light propagation. That is, light sort of "coincidentally" travels at the ultimate speed, but the theory of relativity would apply even it if didn't, so long as this "cosmic speed limit" exists.

A Bit About Waves

The first postulate seems innocuous enough, and perhaps it even seems intuitive. But this business about every observer measuring the same speed for a beam of light didn't sit well with many physicists at the time Einstein proposed it. To see why, and to fully understand the implications of this being true, we need to review a little bit about waves, because as we know from Physics 9B and 9C, light is a wave phenomenon. Of all the things we previously learned about waves, these are the properties of waves we will most need to recall for this discussion...

effect of medium

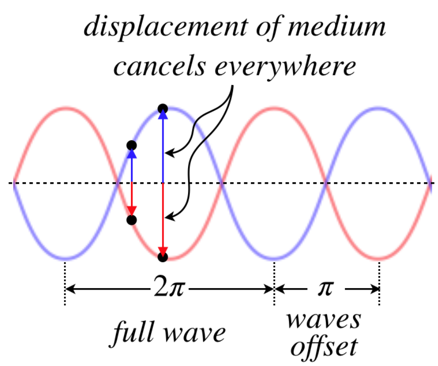
Waves are phenomena which transfer energy through space by the means of a self-propagating disturbance of a medium (the stuff that is "waving"). A wave on a string carries energy along because a piece of the string displaces, which pulls on an adjacent piece of string, displacing it, and so on. Surface waves on water and sound waves through air work in a similar manner. What these all have in common is that medium itself doesn't travel with the energy (it moves with the disturbance), and the speed with which adjacent particles in the medium interact with each other is a property of the medium. Put another way, the speed of a wave is exclusively a function of the properties of the medium. A wave on a string travels faster when the string is more taut, and slower when the string is made is more dense, for example.

superposition

When two waves traveling through the same medium encounter each other, the effects they have on the medium are additive at that the same point in the medium. So if one wave displaces a string by y_1 at some position x on the string, and another wave displaces the string by y_2 at the same position, then the total displacement of the string with the two waves present at x at the same time is y_1+y_2. This additivity property is called *superposition*, and it has the particularly interesting feature that two waves can actually *cancel* each other entirely (a phenomenon called *destructive interference*), if the two waves happen to be displacing the medium equal amounts in opposite directions.

If a wave happens to be harmonic (the medium displacement as a function of position and time is a sine wave), then this destructive interference occurs when two identical waves are out of phase by π.

Figure 1.1.2 – Destructive Interference of Two Harmonic Waves



The Michelson-Morley Experiment

The crux of the problem for those originally opposed to Einstein's assertion that the speed of light is the same when measured by any observer is that *this is not true of other waves*. If one moves through the air into an oncoming sound wave, that sound wave is moving faster relative to that person than relative to someone stationary in the air. The point is that, as stated above, the speed of a wave is entirely determined by the medium, and this speed is *relative to that medium*. So if an observer moves relative to a medium, then the relative speed of waves propagating through that medium can change.

So we are left with the question: What medium is disturbed as a light wave passes through it, and can't we see a change in the speed of light if we just move relative to this medium? This was an open question at the time of Einstein, and was very puzzling for a couple reasons. The first is that the theory of light did not require the existence of any medium at all.

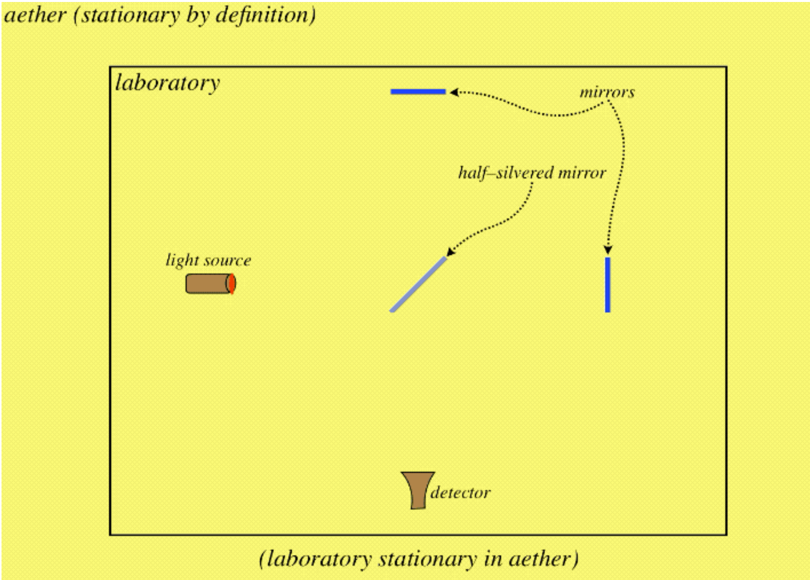
But far more puzzling than this was the fact that no one could seem to come up with any experimental evidence of the existence of a medium for light (which they referred to as the *luminiferous aether*). The most convincing null result of such a search was performed by two American physicists Albert Michelson and Edward Morley, who employed a device called an *interferometer*, the basics of which are still used today for countless applications.

The basics of an interferometer work like this: A single light beam is split into two separate beams. As they came from the same source, they are in phase with each other, but if we send them on separate journeys, and then bring them back together, they may no longer be in phase. For example, one of the beams may travel farther than the other. If the difference in their phases when they get back together is just right, they will destructively interfere with each other.

[*For the two beams to behave this way, the original beam needs to be somewhat coherent, meaning that parts of the light near each other are in phase. Nowadays we easily achieve this artificially with lasers (sunlight is also quite coherent), but fortunately the degree of coherence achievable in a lab in the late 1800's was sufficient to successfully perform this experiment.*]

The ingenuity of the MM interferometer is that it splits the beam of light and sends the two pieces on journeys that are equal distances, but are at right angles to each other. If the laboratory is stationary in the aether, then the beams come back to the detector in phase, and do not cancel each other out:

Figure 1.1.3 – Laboratory Stationary in Aether

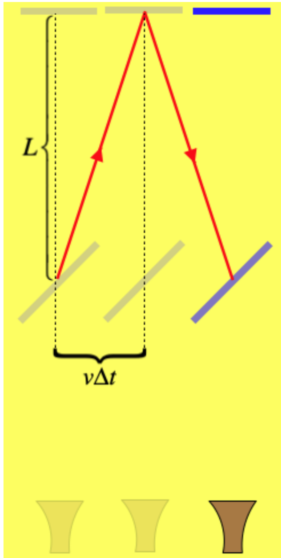


If this device is moving through aether (let's say in a direction parallel to one of these directions), then one of the arms of the interferometer moves parallel to the direction of the light during the journey, while the other moves perpendicular to it. These two journeys are not equal in distance, which means the beams can enter the detector out of phase:

Figure 1.1.4 – Laboratory Moving through Aether

We can do a bit of math to determine the difference in the distance between these two journeys. The only difference between the paths of the two beams comes after they encounter the half-silvered mirror the first time, and before they encounter it the second time, so we focus on this portion of the process for both of them:

Figure 1.1.5 – Distance Traveled by the Transverse Beam



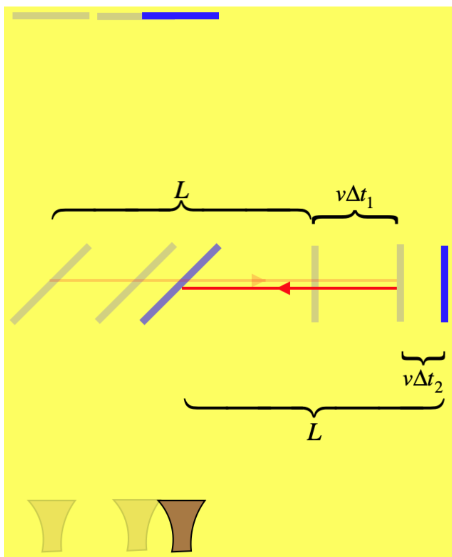
The transverse beam (the one perpendicular to the laboratory's motion) follows a diagonal path in both directions, the distance of which can be determined from the Pythagorean Theorem:

$$s=2\sqrt{L^2+(v\Delta t)^2} \tag{1.1.1}$$

But the distance traveled by the beam also equals the speed of light through the aether c multiplied by the time elapsed, so we get:

$$s=2c\Delta t \quad\Rightarrow\quad s=2\sqrt{L^2+\left(v\frac{s}{2c}\right)^2} \quad\Rightarrow\quad s=\frac{2L}{\sqrt{1-\frac{v^2}{c^2}}} \tag{1.1.2}$$

Figure 1.1.6 – Distance Traveled by the Longitudinal Beam



The longitudinal beam (the one parallel to the laboratory's motion) stays on a straight line, but the time it spends going in each direction is not the same, so we compute these times separately then use them along with the speed of light to get the full distance this beam traverses during this part of the process:

$$\left.\begin{array}{lll} s_1=L+v\Delta t_1=c\Delta t_1 & \Rightarrow & \Delta t_1=\dfrac{L}{c-v} \\ s_2=L-v\Delta t_2=c\Delta t_2 & \Rightarrow & \Delta t_2=\dfrac{L}{c+v} \end{array}\right\} \quad \Delta t_1+\Delta t_2=\frac{2LC}{c^2-v^2} \quad\Rightarrow\quad s=c\left(\Delta t_1+\Delta t_2\right)=\frac{2L}{1-\frac{v^2}{c^2}} \tag{1.1.3}$$

We can see from these two results that the two beams travel different distances. The denominator for each case is a fraction that is less than 1, so the square root is the larger denominator. This means that the longitudinal path (with the smaller denominator) is longer than the transverse path. The difference between these paths can be expressed by a multiplicative factor that depends upon the speed of the laboratory through the aether and the speed of light:

$$s_{longitudinal} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} s_{transverse} \tag{1.1.4}$$

We therefore find that if the laboratory is moving through the aether at just the right speed, then the difference in distances traveled will equal half the length of a wave, and the two beams will cancel at the detector, resulting in darkness. In fact, this is not the only case that will give this result. If the difference in distance traveled is $\frac{3}{2}$ of a full wave, then again the troughs and crests of the waves will match up, making destructive interference. Moreover, even if there is not complete destructive interference, every offset of the waves will result in some variation in the brightness of the light.

So given that the earth is moving through space, and is also rotating, one would expect that it is not difficult to find some evidence that the aether exists, and that we are moving through it. But try as they might, Michelson and Morley found no such thing. Many explanations were offered for the problem (many before the experiment was even performed), such as the aether being "dragged" by the earth, so that it was stationary around us here on the surface (logical arguments based on starlight observations and additional experiments on starlight proved this to be false). Einstein alone showed the courage to just discard the existence of the aether altogether, no matter how nonsensical the consequences might seem to be at first.

Analyze This

A

Your goal in the analysis is to extract everything you can from what has been given. At a minimum, every analysis should include these items:

- *what we are given (perhaps translated from English to mathematics)*
- *what we can infer, if anything*
- *quantities we can compute (or almost compute!), if anything*

Analysis

F

1.2: The Nature of Time

Spacetime Events

We now embark on deriving the consequences of the relativity principle in the same way that Einstein did – using a tool he called *Gedankenexperiment* (*thought experiment*). in order to keep everything straight in our discussions, we begin by defining a *spacetime event*.

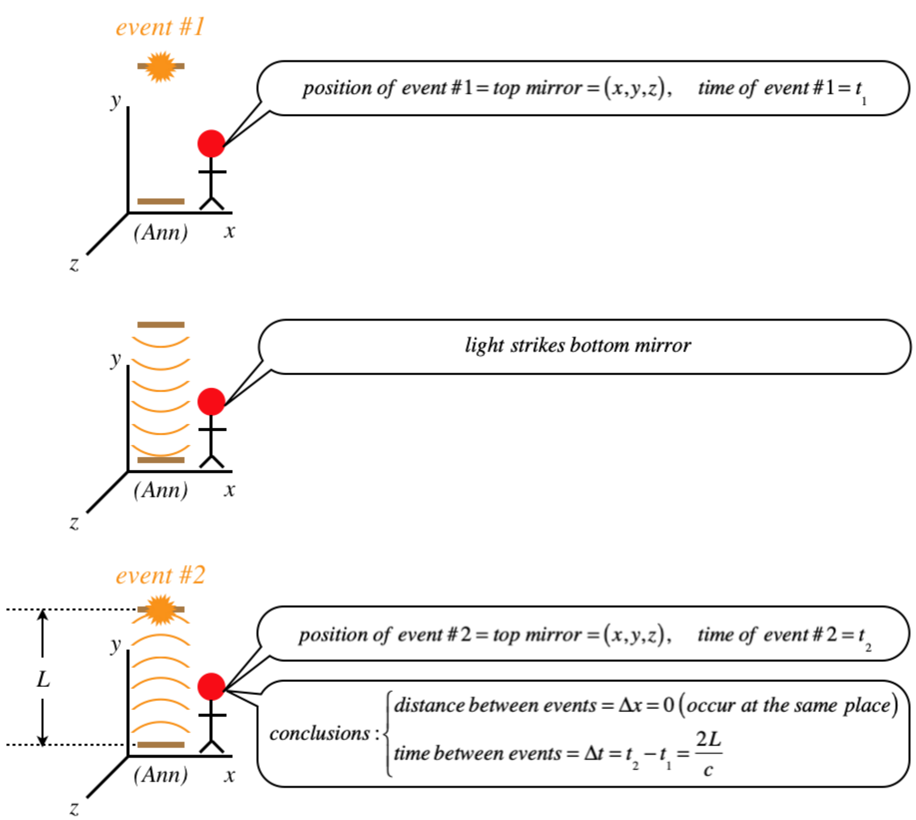
In the context of special relativity, a spacetime event is an instantaneous occurrence at a specific point in space and at a specific moment in time. A single point on a stationary light bulb as it dims defines a specified location, but it is not an event because the dimming process does not occur at a single instant in time. A baseball bat at exactly 12:01pm occurs at a single instant in time, but it is not an event, because the position is not specified at a single point. An easy way to visualize a spacetime event is to picture it as a very quick flash of light from a point source. The position of the point source and the instant in time the flash occurs define the space and time "coordinates" of the spacetime event.

It is much easier to define what a spacetime event is than it is to put physical quantities in terms of the spacetime coordinates, but as we will see, this is exactly what we will have to do to make sense of what is to come. We begin with one of most startling results, which is ironically one of the easiest to derive.

Time Dilation

Our first thought experiment involves turning the function of a clock into a series of spacetime events. This clock functions as follows: Light bounces back-and-forth between two mirrors, and every time it strikes one of the mirrors, the clock "ticks." We begin with Ann's perspective on what is happening with this clock. She happens to be in the same frame as the two mirrors, so to her they are at rest, and the light is bouncing parallel to her y-axis. The two spacetime events we will look at are two consecutive ticks of the clock.

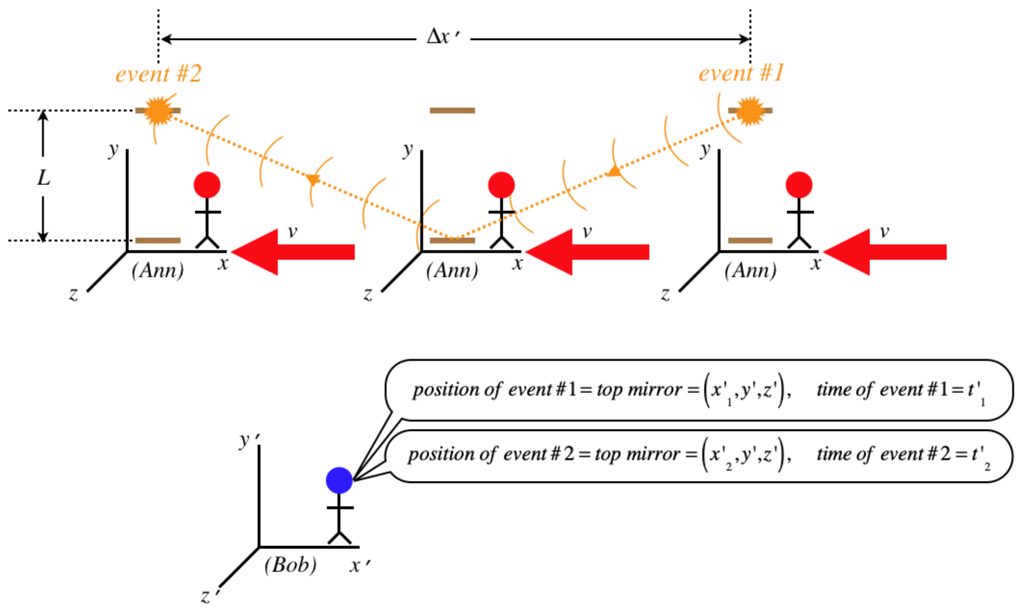
Figure 1.2.1 – Ann's Perspective of the Light Clock



Okay, so we have used the two events to determine the time span between them according to Ann. The goal of relativity is to describe what a second observer measures for a physical process given what the first observer measured. So now we introduce Bob, who is in what we call the primed inertial frame, moving at a constant speed v in the $+x$-direction relative to Ann. One might interject, "Wait, this is *time* we are talking about! Won't both of them measure the same amount of time between ticks of the clock?" Don't assume anything in relativity – just use the spacetime events and the postulate(s), and see where it leads.

Looking from Bob's perspective means that not only is Ann moving in the $-x$-direction, as we noted previously, but the two events (which both occur at the top mirror) don't occur at the same position in space, since the mirror moves:

Figure 1.2.2 – Bob's Perspective of the Light Clock



Now we calculate the time between the two events, as we did for Ann. From Bob's perspective, the light travels a longer distance than Ann measures, and very importantly, *both Ann and Bob measure the speed of light to be the same* (postulate of relativity), so Bob must measure a longer time period than Ann measures between the same ticks of the light clock! According to Bob, the light travels diagonally from the top mirror to the bottom one, and the length of this half of the trip can be written in terms of the speed of light, and in terms of the Pythagorean Theorem:

$$\begin{aligned} \Delta x' &= x'_2 - x'_1 = v\Delta t' \\ c\Delta t' &= 2\sqrt{L^2 + \left(\frac{\Delta x'}{2}\right)^2} \end{aligned} \tag{1.2.1}$$

We can eliminate $\Delta x'$ from these two equations to relate the time span measured by Bob to the time span measured by Ann:

$$\Delta t' = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\left(\frac{2L}{c}\right) = \gamma_v \Delta t\ , \quad \gamma_v \equiv \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{1.2.2}$$

The time between ticks for Bob is greater than the time between ticks for Ann by a factor of γ_v (which is clearly a constant greater than 1). Just to clarify, this is not an optical illusion for Bob – he doesn't just "see the clock ticking slower than it really is," *it is actually ticking slower*. Also, it is important to note that while we used light to achieve this answer, it doesn't just apply to light phenomena, it applies to time flow in all its manifestations. If Ann measures her own pulse to be 60 beats per second (one second between each beat), and $\gamma_v = 2$, the Bob would measure Ann's heart rate to be 30 beats per second (2 seconds between each beat).

It's worth taking a moment to review what the source of this result is. It comes from the fact that the light in the light clock travels farther for Bob than it does for Ann, but they agree on the speed of that light, which means that the time between the two events must be greater for Bob than it is for Ann.

As startling as this result is, it gets weirder. Suppose Bob has a light clock exactly like Ann's. What does Ann observe when she looks at Bob's clock? She sees exactly the same thing happening with Bob's clock as he sees with her clock! Therefore Ann claims that time is passing slower for Bob than it is for her, even as Bob says that Ann's time is passing slower than his own. Which one of them is correct? Is Ann's time passing slower, or is Bob's? They are both in inertial frames, so according to the principle of relativity, each has an equal right to declare themselves to be "stationary." Therefore they are both right. The reason it seems like it is impossible that this can be true is that we cling to the incorrect notion that time is universal. The time span between two events is a relative quantity that depends upon who measures it.

Recording Spacetime Coordinates

While the calculation above is correct, it does require an assumption that we need to briefly address. Both Ann and Bob noted the positions and times of the events in their frames. Given the importance of both position and time in relativity, we need to be specific about how these numbers are recorded. What is observed is a spacetime event, which we have modeled as a flash of light that occurs in an instant at a specific position. So let's imagine constructing a massive lattice of labeled positions throughout all of space, and the position of any possible flash must coincide with one of those positions, giving us our spatial label. Note that every inertial observer can create such a lattice independent of every other observer, because according to the relativity principle, everyone has an equally valid claim to being "stationary." It is true that Bob's lattice of position labels is moving according to Ann, but Ann and Bob only use their own stationary labels to describe the positions of events they see.

To get a complete reckoning of an event, we need to record not only its position, but the time at which it occurs. Given what we know about the rate of time flow for moving clocks, we have to be very careful about how we measure the time at which an event occurs. The one way to be safe is to have the clock that reads the time be positioned at the same place in space as the event. So whenever an event occurs, one simply reads the label of the lattice point at which it occurs, and the value indicated by a clock located at that lattice point when the event occurs.

Two Different Time Measurements

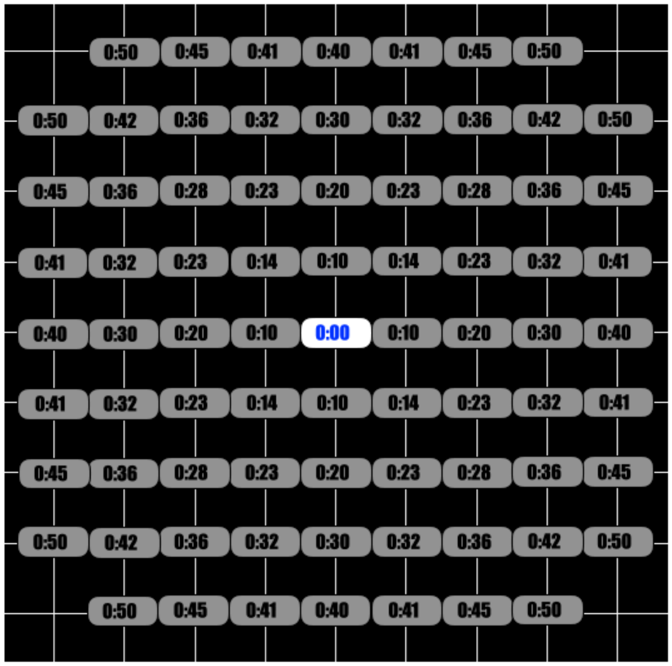
Now that we have a plan for recording data for events in spacetime, we need to give a little more thought to how we plan to have a clock that is properly positioned to measure the time. It turns out that there are two fundamentally different ways to achieve this.

coordinate time

The first way that comes to mind for measuring the time of any given event is to simply place a separate clock at every lattice point. While this is a simple way to get a measurement for any event, we will be interested the time intervals *between two events*, which means that all of our clocks positioned throughout space need to be synchronized. How do we do this? If we bring all our clocks together in one place, set them the same, and then move them out to their assigned locations, then the weird effects that come from relative motion of clocks make un-synchronize them when they are moved. Instead what we can do is this

1. Distribute all of the clocks throughout space.
2. Set the clock at the origin to a time of 0:00.
3. Using the lattice positions of the clocks, compute the time it will take a spherical wave pulse of light that starts at the origin to reach all the other clocks, add this time to 0:00, and set the clock at this time.
4. Start the clock at the origin while starting the spherical pulse of light from the origin.
5. When the light wave reaches a clock, start it running.

Figure 1.2.3 – Synchronizing Clocks Distributed in Space



By anticipating what the time on the origin clock must be when the light arrives, we can assure that the spherical wave propagates clock synchronization throughout space. This measurement of the time of an event is called *coordinate time t*.

In the example above, both Ann and Bob measured the time between ticks in their own coordinate time. For Ann, the coordinate time span between the two spacetime events was $\Delta t = t_2 - t_1$, while for Bob it was $\Delta t' = t'_2 - t'_1$. As we found in the thought experiment, these values are not equal, which is to say that this manner of measuring a time span between two events is *relative*. Whenever the value of a physical quantity is different when measured from different inertial frames, we say that such a quantity is *frame-dependent*. We therefore declare:

Coordinate time spans are frame-dependent.

proper time

We certainly are not *required* to measure time between events by placing synchronized clocks at all the lattice points in our frame. Another way would be to use a *single* clock that is moved from the lattice point of the first event to the lattice point of the second. As before, a clock records the time of the event while it is at the same point in the lattice as the event, but this time it is the *same* clock, which means we do not need to rely upon our synchronization method above. A time interval measured in this manner is called a *proper time $\Delta\tau$* between the spacetime events.

Alert

The name "proper time" dates back to the early days of relativity, and is still used today, but it is dangerously misleading for those new to the subject. The word "proper" can easily be misconstrued to mean "correct," and hopefully this section is making it clear that this is cannot be the case. We are in the process of defining two different ways of measuring the time between two events (which can give different answers), and neither of these is any more correct than the other. The sooner the reader purges from their thoughts the notion that time is absolute and that there must be one correct value for the time between two events, the better.

It might seem like both the coordinate and proper time methods of measuring time intervals between events should produce the same result, but in fact they do not. The figures below demonstrates these two measurements for the same two spacetime events.

Figure 1.2.4 – Spacetime Events

Here we have just two spacetime events viewed from a particular reference frame.

Figure 1.2.5 – Coordinate Time Interval

These are the same two events, viewed from the same reference frame, but the coordinate time clocks placed at the positions of the events are in place to measure the times at which the events occur.

Figure 1.2.6 – Proper Time Interval

Again we have the same two spacetime events, viewed in the same frame, but a clock at rest in a different frame is now visible, and it measures the time interval between the two spacetime events, *but in that frame the events occur at the same position* (at the nose of the rocket ship).

Our example with Ann and Bob earlier shows why these time measurements come out different, if the postulate of special relativity about the constancy of the speed of light is accurate. The two flashes occur at the same lattice point in Ann's frame (the top mirror remains at the same place in Ann's labeled lattice), so she measures the time interval using the proper time method. Meanwhile, the top mirror moves from one lattice point in Bob's frame to another, so he relies upon the synchronized clocks positioned at those points for the time interval. The thought experiment demonstrates through the postulates of relativity that these two time intervals are not equal.

Example 1.2.1

In the figures above that depict two spacetime events and a spaceship moving between them, we are observing from an inertial frame. How fast is the spaceship moving relative to this frame?

Solution

The two events occur at the same position in space in the ship's frame (at its nose - remember that the people on the ship can claim that the ship is not moving, so the two flashes occur at the same place). Therefore the ship measures the proper time interval between the two flashes, just as Ann measured the proper time interval of the two flashes at the top mirror. Using the formula we derived to express the relationship between the time intervals, we find:

$$\Delta t=\gamma_v\Delta\tau \quad\Rightarrow\quad \sqrt{1-\frac{v^2}{c^2}}=\frac{\Delta\tau}{\Delta t}=\frac{12s}{15s}=0.8 \quad\Rightarrow\quad v=0.6c$$

The feature that best distinguishes proper time from coordinate time is the fact that a coordinate system is not needed to measure proper time. For example, we could introduce several other inertial frames of reference to look at the time interval between those top mirror flashes, but *only* Ann's clock will measure the time interval as the the one where the flashes occur at the same position in space – every frame other than Ann's will be similar to Bob's, in that the flashes occur at different lattice positions for their frame. All of the observers will agree on one thing – that Ann's measurement of the time interval is somehow "special", and this gives them a way to all agree upon a time interval. Put another way, the measurement of proper time (essentially asking Ann what answer she got, as she was the only observer in an inertial frame to have both events occur at the same position) gives the same result for all reference frames. That is:

Proper time is frame-independent.

Besides "frame-independent," a word typically used to describe a physical quantity like proper time that doesn't vary from one frame to another, is *invariant*.

Note that it is possible for a proper time measurement to be equal to a coordinate time measurement. For example, in the case discussed above, Ann sees the two events occur at the same lattice point in her frame, so if she looks at the clock placed there, it is the same clock measuring the time for both events, which means it also records the proper time. Bob's measurement of coordinate time, on the other hand, is not the proper time, since he reads the numbers off two different clocks – one placed at the lattice point of the first spacetime event, and one placed at the second. From the light clock example, it should be clear that the shortest distance the light has to travel between the two mirrors occurs in Ann's frame. That is, *every* frame other than the "proper frame" that measures coordinate time is going to measure a longer time interval between the events than the proper time.

two important notes

There are two details that have not been tied-up above that we will mention here and address in a future section:

1. The proper time between two events may not be definable as a real number, if the events are separated by a distance that is too great for the spaceship to cross in the interval between their occurrences. For example, if the two events viewed in the frame of the figure above occurred simultaneously in that frame, then there is no way for the spaceship to traverse the distance between the events fast enough to allow both events to both occur at the nose of the ship. We will see later that there is another invariant quantity that applies to any two events, and that the proper time interval is a special case of this invariant when an inertial frame exists that can measure the events at the same position in space.
2. We have defined the proper time here as being measured in an inertial frame, but the spaceship could also have both events occur at its nose when it accelerates between the two events. This will come out to a different result than the inertial frame case, so it is important when declaring (as we did above) that the "proper time is frame-independent", that we keep in mind that this is restricted to the specific history of the clock that measures it. That is, every observer will agree on the proper time measured by a single clock that is present at the positions of both events. "Invariance" pertains to different observers, *not* different clocks. A second clock that is present at the positions of both events will not necessarily measure the same proper time interval as the first clock. The *way* that each clock gets from the first event to the second (namely, its acceleration during the trip) determines how these two proper time measurements may differ.

Cosmic Speed Limit and the Spacetime Interval

When we look back at the time dilation result we obtained above, an obvious question comes to mind: If we observe a clock in a moving frame to tick more slowly than one in our rest frame by a factor of $\sqrt{1-\frac{v^2}{c^2}}$, then what happens when the relative speed of the two inertial frames reaches or exceeds $v=c$? Clearly the result gives a nonsense answer, and while this is far from "proof," we will take this moment to make a declaration that we will later see to be true in many other cases...

Two inertial frames can never have a relative speed that exceeds the speed of light, and this cosmic speed limit can only be attained for light itself.

Technically, there are other phenomena besides light that can propagate at light's eponymous speed, and the criterion for this is a simple one, but we will save that discussion for later. For now, we will generalize the "cosmic speed limit" to state that no "influence" or "information" can be passed from one point in space to another at a speed faster than light can traverse the same distance.

This speed limit gives us a new perspective on time intervals between two events. We said above that there was no way to measure the proper time interval between events that are separated in space and are simultaneous, because there is no way for a single clock to get from one event to the other in time. Now we see that the two events don't need to be simultaneous for this to be true. Because of the cosmic speed limit, there is no way for the proper time interval between two events to be measured if they are separated by enough distance that light cannot travel from the earlier spacetime event to the later one. If light can't make it in time, then neither can a clock, and the proper time cannot be measured.

Let's say that the events occur at positions (x_1,y_1,z_1) and (x_2,y_2,z_2), and times t_1 and t_2 (with $t_2>t_1$), as measured in some arbitrary inertial frame, respectively. Then there will be a well-defined proper time measurable by a moving clock if the distance between them is less than the distance that light can travel in the time interval $\Delta t = t_2 - t_1$:

$$\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} < c\Delta t \tag{1.2.3}$$

We can therefore invent a sort of "discriminant" that tells us whether two events can be connected in this way:

$$\Delta s^2 \equiv c^2 \Delta t^2 - \left(\Delta x^2 + \Delta y^2 + \Delta z^2\right) \tag{1.2.4}$$

When $\Delta s^2 > 0$, it is possible to move a clock from the earlier event to the later one, so that the clock measures the proper time, otherwise one cannot do this. This quantity Δs^2 is called the *spacetime interval* between the two events.

We said that the quantities Δt, Δx, Δy, and Δz are measured in any arbitrary inertial reference frame, but for a moment let's suppose that the events are sufficiently close together, and look at the value of the spacetime interval in the frame where the events are at the same position (i.e. the frame with the clock that, when viewed by someone else, is moved from the earlier event to the later one). In this frame, $\Delta x = \Delta y = \Delta z = 0$ and Δt is the proper time, which means:

$$\Delta s = c\Delta\tau \tag{1.2.5}$$

So the spacetime interval between two events is just proportional to the square of the proper time interval between those events. We stated earlier that the value of $\Delta\tau$ is an invariant – it is the same when measured in any reference frame. This means that all observers will agree on the spacetime interval between two events that allow for a proper time. But now that we are talking about an abstract mathematical quantity instead of a span of time, we can make the more general statement for *all* pairs of events:

The spacetime interval between any two events is an invariant.

Yes, sometimes this interval is positive (allowing for a measurable proper time interval), sometimes it is negative (not allowing this - the proper time interval is imaginary, whatever that means), and sometimes it is zero (making the proper time interval zero). But whatever it comes out to, the value of Δs^2 is measured to be the same quantity in all frames of reference. Let's be clear about what this means: The values of Δt, Δx, Δy, and Δz are all different for various reference frames, but the special combination of these quantities that equals Δs^2 comes out to be the same in every frame, provided the same two events are involved.

There is one last observation we should make about the spacetime interval between two events. Suppose we consider two events that are separated by infinitesimal differences in distance and time. Then we have:

$$ds^2 = c^2 dt^2 - \left(dx^2 + dy^2 + dz^2\right) \quad \Rightarrow \quad ds = dt\sqrt{c^2 - \left(\left[\frac{dx}{dt}\right]^2 + \left[\frac{dy}{dt}\right]^2 + \left[\frac{dz}{dt}\right]^2\right)} \tag{1.2.6}$$

Given that the time between the infinitesimally-separated events is Δt and the distance in the x-direction between these events is Δx, then the ratio $\frac{\Delta x}{\Delta t}$ is the speed that the clock must move (measured in our arbitrary frame) in the x-direction to get from the earlier event to the later one to record both times. The same is true for the y and z directions, so:

$$ds = dt\sqrt{c^2 - \left(v_x^2 + v_y^2 + v_z^2\right)} = cdt\sqrt{1 - \frac{v^2}{c^2}} \tag{1.2.7}$$

One can now imagine computing (the square root of) the spacetime interval between two events of finite separation by "chaining together" (integrating) infinitesimal intervals:

$$\Delta s = \int\limits_{event\ A}^{event\ B} ds = \int\limits_{A}^{B} cdt\sqrt{1 - \frac{v^2}{c^2}} \tag{1.2.8}$$

Combining this with Equation 1.2.5 gives us the link between the proper time interval and the coordinate time interval between events A and B:

$$\Delta\tau = \int\limits_{A}^{B} dt\sqrt{1 - \frac{v^2}{c^2}} \tag{1.2.9}$$

Where v is the relative speed of the coordinate frame and the "proper frame" (the frame where the two events occur at the same place). To avoid complications, we will assume that the path between events is a straight one, but in general it could involve constant speed or accelerated motion. If the speed is constant, then the integral is trivial and we get the same result as Equation 1.2.2 – the inertial frame time dilation formula we found from our thought experiment. But if the speed is not constant, then the quantity in the square root factors into the integral, and the proper time is not the same as it was for the case of the inertial frame. Later we will see the important consequences of this.

1.3: More Thought Experiments

Doppler Effect

We know that two observers in motion relative to each other measure the same speed when they look at a light wave, but what about the other properties of the light, such as the frequency and wavelength? We know that for sound they would not measure the same frequency due to the *doppler effect* (the phenomenon responsible for the change in perceived pitch of a car siren as it drives by), but in the case of the doppler effect for sound the medium through which the sound travels plays a critical role. In particular, an observer moving through the air toward a sound source will note that the sound wave is moving toward them faster than that sound is moving through the air. This of course is not the case for light traveling through a vacuum. Nevertheless, there *is* a doppler effect for light. To get started, we need to define a few things about waves:

The length of the repeating waveform, called the *wavelength* of the wave, we represent with the symbol \(\lambda\. A snapshot of the wave tells us something about its spatial features like the wavelength and amplitude, but the wave is moving, so if we want to know something about its time-dependence, we need to select a specific point in space, and observe the displacement of the medium (or in the case of light, where no medium is needed, the strength of something called the electromagnetic field – but more on this in classes yet to come). The wave moves at a constant speed, and the length of each repeating waveform is the same, so the time span required for a single waveform to go by is a constant for the entire wave, called the *period* of the wave. An alternative way of measuring the temporal feature of the wave is the rate at which the process repeats, called *frequency*. Frequency is measured in units of cycles per second, a unit known as *hertz* (*Hz*). Since 1 period is the time required for one cycle, there is a simple relationship between period and frequency:

$$f=\frac{1}{T} \tag{1.3.1}$$

We can make another association of periodic wave properties. If we pick a specific point on a waveform (called a *point of fixed phase* for the wave), and follow its motion, it should be clear that it travels a full wavelength in the time of one period. We therefore can relate the wave speed, wavelength, and period (or frequency):

$$c=\frac{\lambda}{T}=\lambda f \tag{1.3.2}$$

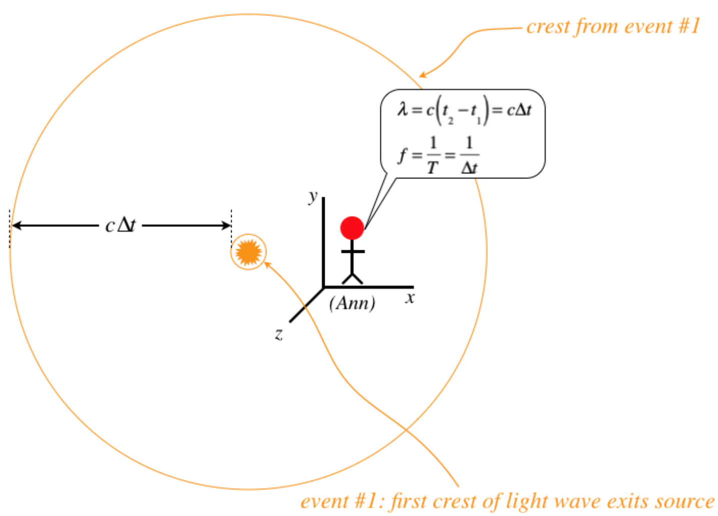
In the analysis to come, we will represent the "crests" of light waves with circles, so that the distance between these circles is the wavelength. We'll start with the basic phenomenon of doppler effect. The two gifs that follow apply equally to light or sound. The main idea is to note that when there is no relative motion, the rate of flashes of the red source equals the rate of flashes of the blue receiver (each flash of the receiver occurs when a crest arrives). But when the source is moving relative to the receiver, the rate of source and receiver flashes do not match. Specifically, the receiver frequency goes up when the relative motion is toward each other, and goes down when it is away from each other.

Figure 1.3.1 – No Relative Motion of Source and Receiver

Figure 1.3.2 – Approaching Relative Motion (Receiver's Perspective)

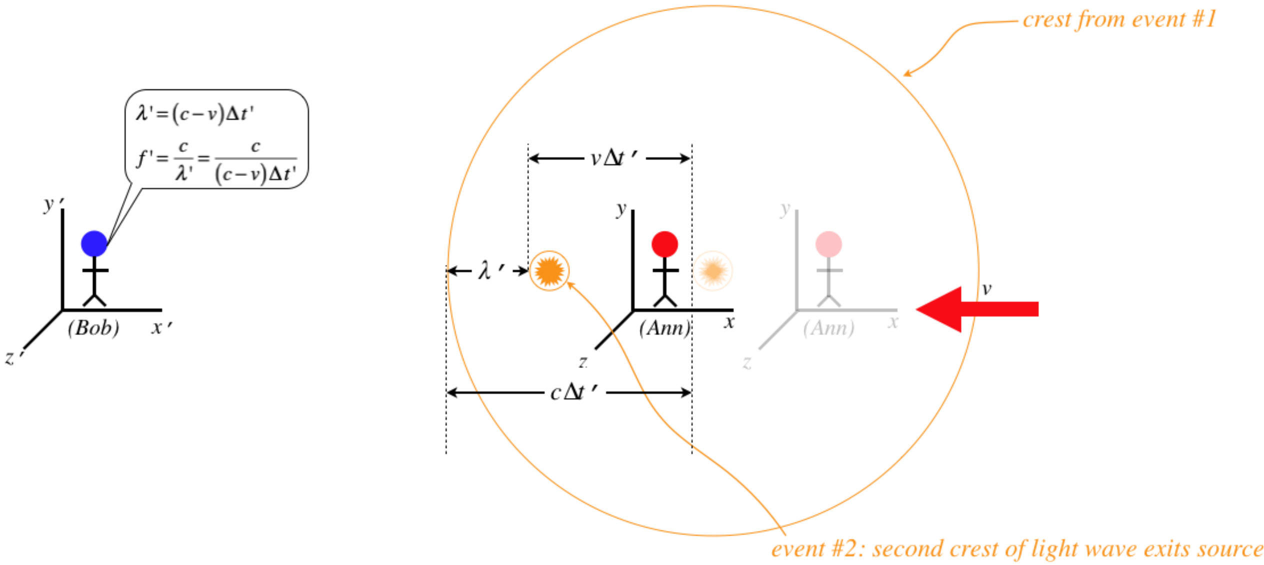
Now let's put Ann and Bob into the roles of source and receiver, respectively. We start with what Ann measures for the light source that remains stationary in her frame:

Figure 1.3.3 – Ann's Perspective of Light Signal for Two Crests



The wavelength's relationship to the frequency and wave speed is as we stated above.

Figure 1.3.4 – Bob's Perspective of Light Signal for Two Crests



Bob's measurement of the wavelength is different from Ann's because in the time between emissions of the two wave fronts, the source moves according to Bob. But this is not the only source of the disparity between the two frequency measurements. The diagram shows that the time elapsed between wave front emissions also plays a role, and as we know, the fact that the flashes occur at the same place for Ann means that her coordinate time measurement happens to equal the proper time measurement (and she is in an interial frame, so this is also the spacetime interval), while Bob's coordinate time measurements involve events at different positions. So Bob will measure a longer time between flashes than Ann. While all waves (most notably sound) exhibit the doppler effect, the result is different for light, giving us an "ordinary" doppler effect, and a relativistic doppler effect. Plugging in the time dilation relation gives the relation between the two frequencies measured:

$$f'=\frac{c}{(c-v)\,\Delta t'}=\left(\frac{c}{c-v'}\right)\left(\frac{1}{\gamma_v\Delta t}\right)=\left(\frac{c}{c-v}\right)\left(\sqrt{1-\frac{v^2}{c^2}}\,\frac{1}{\Delta t}\right)=\sqrt{\frac{c+v}{c-v}}\;f(\text{moving toward each other}) \tag{1.3.3}$$

Whenever this occurs with light in the visible spectrum, the change in frequency goes away from the red end of the spectrum, and toward the blue end, so this increase in frequency for light is called a *blue shift* (even when the light is not in the visible spectrum).

If Ann happens to be moving *away* from Bob, then it is a simple change to this equation to get the correct answer – change v to $-v$, giving:

$$f'=\sqrt{\frac{c-v}{c+v}}\;f(\text{moving away from each other}) \tag{1.3.4}$$

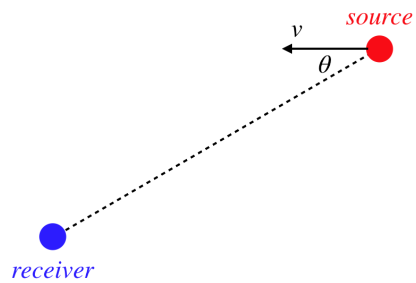
This effect of reducing the frequency perceived due to relative motion is called a *red shift*.

But of course these are not the only two options. For example, Ann and Bob may be in the process of moving past each other. If this is happening, then determining the wavelength measured by Bob is a tougher, but the time dilation between the two frames still applies. One specific example we can look at is when Bob and Ann are aligned along the y-axis. We have to be careful about using words like "when" in the context of relativity, so we will define this moment for Bob as when he *sees* the source of the light as being aligned with him along the y-axis (of course, he will *deduce* that the light source is elsewhere, but that is not what counts here, as Bob is actually observing the light). At this moment, Ann and Bob will agree upon the wavelength, since their relative motion is along the x-direction and has no effect on the spacing of wave fronts. In this case, *only* the time dilation plays a role, and the result is simply:

$$f'=\frac{1}{\Delta t'}=\frac{1}{\gamma_v\Delta t}=\frac{\sqrt{1-\frac{v^2}{c^2}}}{\Delta t}=\sqrt{1-\frac{v^2}{c^2}}\;f \tag{1.3.5}$$

The more general case involves the line joining the source and receiver forming an angle θ with the relative velocity vector:

Figure 1.3.5 – Relative Motion of Source and Receiver Not Along Line Joining Them



Once again, the line joins the receiver and the apparent (not deduced) source of the light. In this case, the doppler effect on the frequency comes from the component of the source's motion relative to the receiver that lies along the line joining them. That is we replace the v above with $v\cos\theta$:

$$f' = \frac{c}{(c - v\cos\theta)\,\Delta t'} = \left(\frac{\sqrt{c^2 - v^2}}{c - v\cos\theta}\right) f \tag{1.3.6}$$

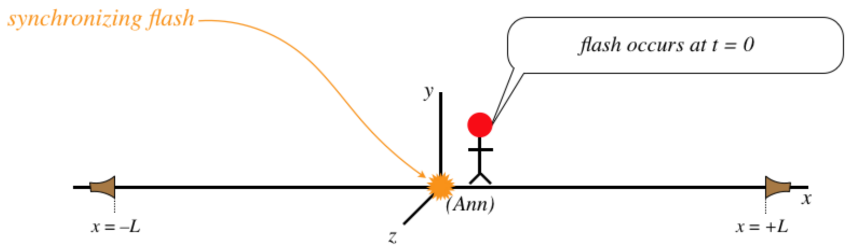
It is left as an exercise for the reader to show that this equation reduces to the three equations above for the appropriate values of θ.

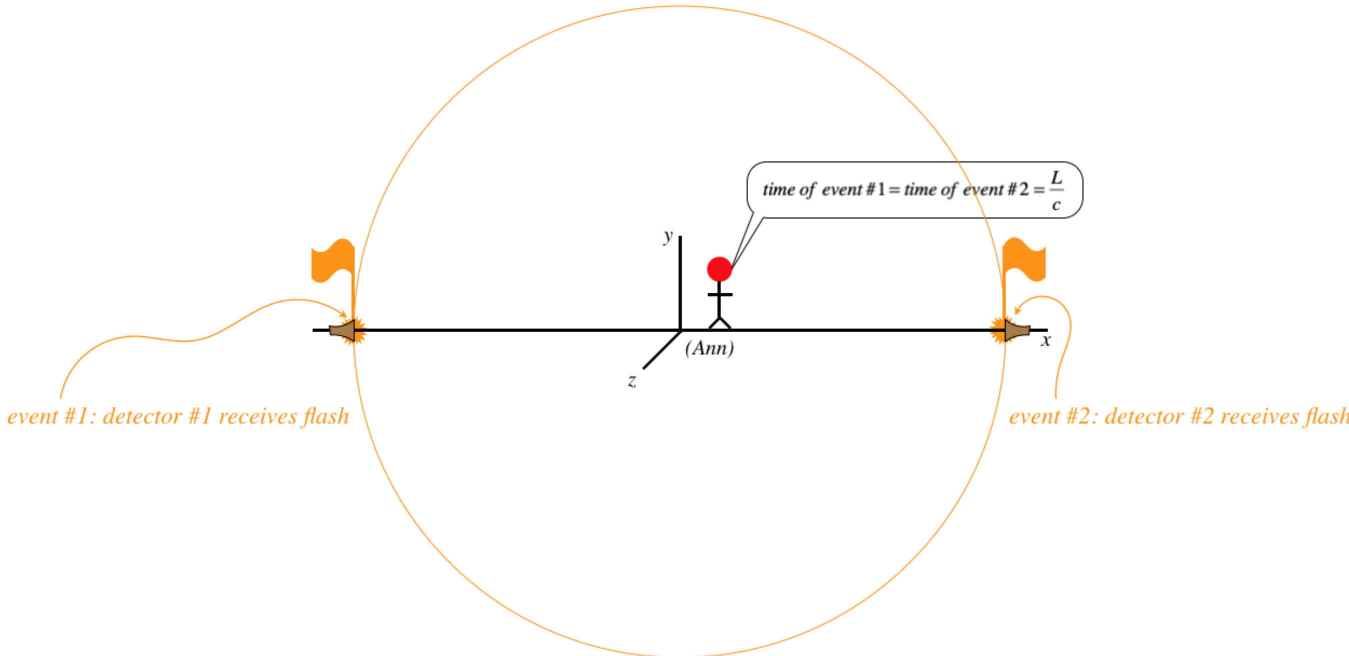
Simultaneity

Let's return to our discussion of how to measure coordinate time by synchronizing clocks at all the lattice points in a reference frame. Suppose Ann and Bob are moving past each other along the x-axis, and at the moment that their origins coincide, they start their clocks at the origin. Then each of them synchronizes all the clocks on their lattice with the clock at the origin. Doesn't this mean that all of Ann's clocks are synchronized with all of Bob's clocks? And if so, doesn't this mean that they should measure the same coordinate times between events, in contradiction to everything we have said fo far? Such a conundrum calls for a thought experiment!

Let's suppose Ann decides to synchronize two clocks using a flash from her clock at the origin:

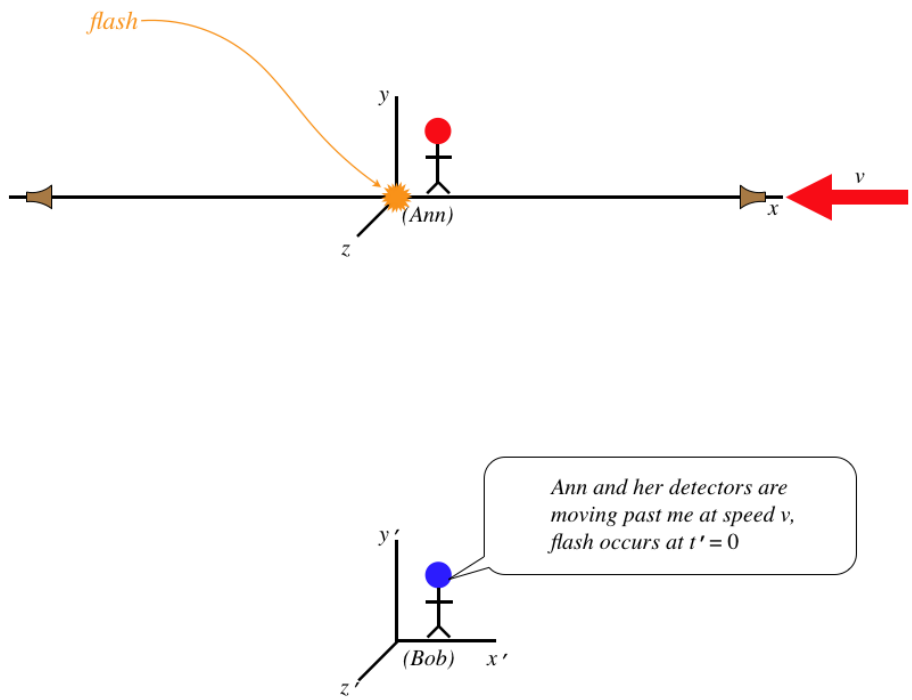
Figure 1.3.6 – Simultaneous Events for Ann





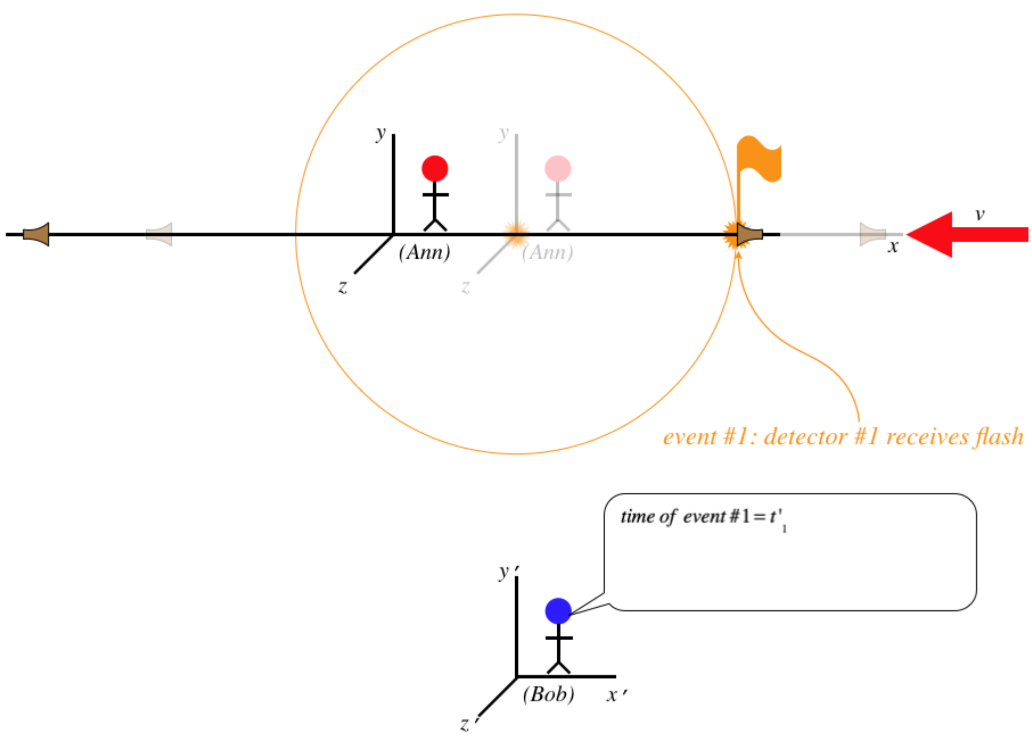
Does Bob agree that the two clocks (located at the two detectors) are synchronized? Let's look at what Bob sees:

Figure 1.3.7 – Ann's Synchronized Events Seen by Bob (a)



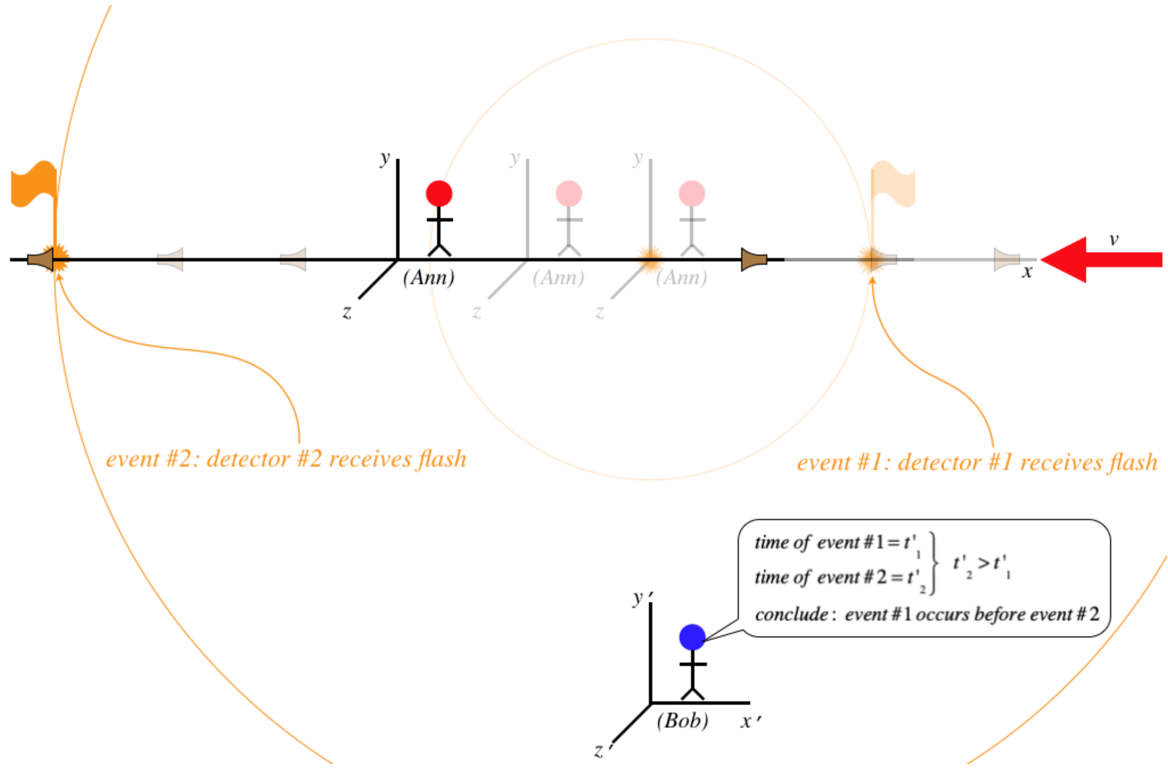
The detectors Ann is using are fixed in the lattice points in her frame, so they move along with her, according to Bob. When the light flashes, it takes time for the wave to get to the detectors, and while this time passes, the detectors move, according to Bob:

Figure 1.3.8 – Ann's Synchronized Events Seen by Bob (b)



As you can see, the detector trailing Ann receives the signal before the other detector, according to Bob.

Figure 1.3.9 – Ann's Synchronized Events Seen by Bob (c)



Far from seeing the two events simultaneously, Bob measures a time difference between them. This means that when he looks at all the clocks in Ann's lattice, he sees them all out of sync, with the times getting later the farther the clock is on the positive side of the origin. We therefore find that the concept of simultaneous events is relative (frame-dependent).

Alert

*It is important to keep in mind that when we are talking simultaneous events in one frame, we are not talking about about simply **seeing** two things occur at a different time. For example, if Ann happened to be standing close to detector #1, then the light from the flag that pops up there would reach her sooner than the light coming from the flag at detector #2, and she would witness the two flags popping at different times, but the two events would still be simultaneous in her frame.*

Example 1.3.1

We found in the light clock thought experiment that the relationship between Ann's and Bob's time measurements is given by Equation 1.2.2. If Ann's two clocks are synchronized, then the time between the two events that occur when the flash reaches both detectors is zero. So why don't we find that for those same two events viewed by Bob, the time interval is also zero?

$$\Delta t' = \gamma_v \Delta t = 0$$

Solution

The equation quoted assumed that the time measured by Ann was the proper time, since the two events occurred at the same position in her inertial frame. The synchronized events in this case do not occur at the same position, so it is the coordinate time that she measures to be zero. One way to avoid this confusion is to write the time dilation formula of Equation 1.2.2 explicitly in terms of the proper time:

$$\Delta t' = \gamma_v \Delta \tau$$

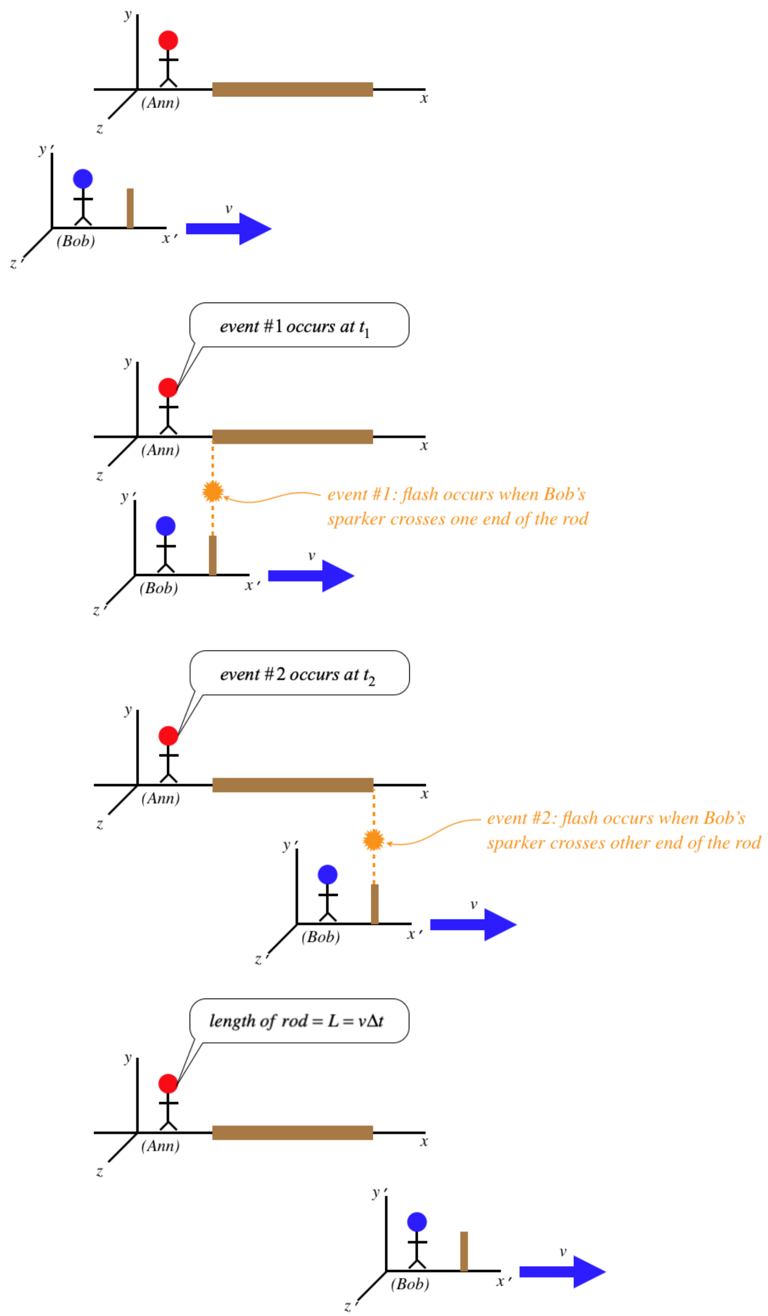
In the case above, the comparison is not between a coordinate time and a proper time, but two different coordinate times.

Length Contraction

Instead of comparing time spans between two frames in relative motion, let's compare distance spans. To do this, we need to first figure out *what it means* to measure the length of an object (say a meter stick). As we know, whatever we do in relativity must be in terms of spacetime events. We can't simply say that the length of an object is the distance between events that occur at the object's endpoints, because the object might move after one event occurs and before the second one occurs. So clearly to define the length of an object, we need to stipulate that the two events that occur at the endpoints of the object being measured occur *at the same time*. But since observers in two frames in relative motion will not agree to what events are simultaneous, it stands to reason that they might not agree to length measurements.

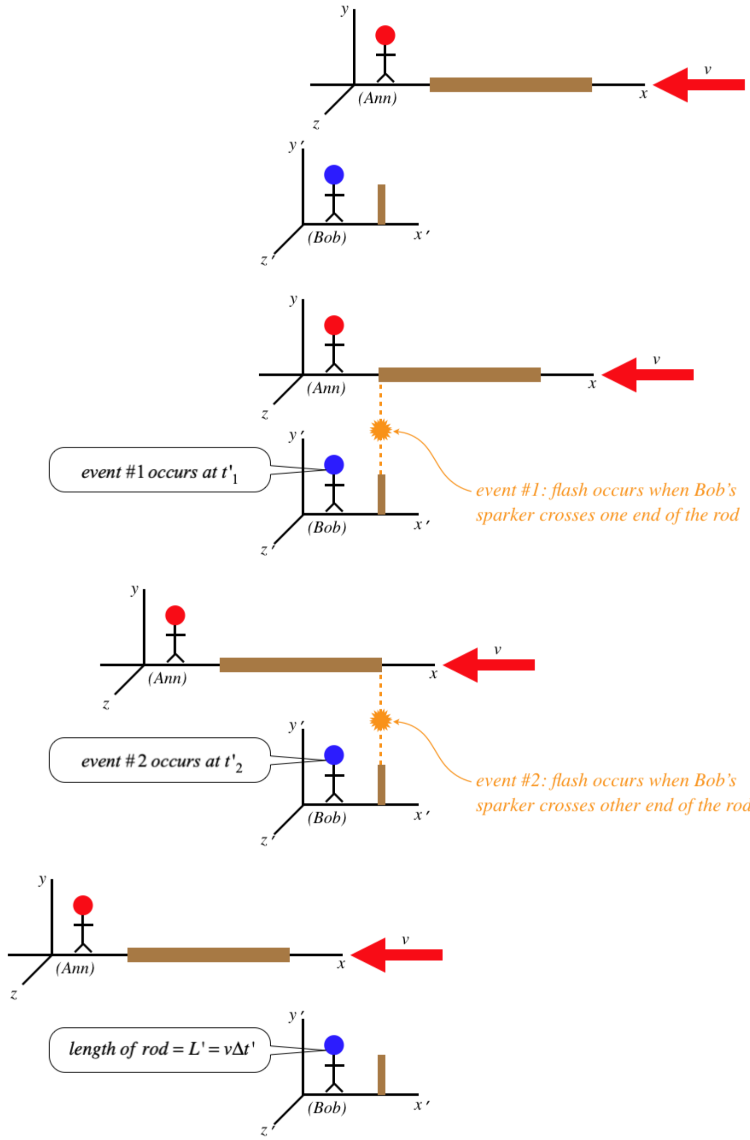
We consider The following scenario: Ann lays a rod down along her x-axis, and as she passes by Bob, each end of the rod creates a spark (constituting a spacetime event) when it coincides with a device stationary in Bob's frame that we will call a "sparker."

Figure 1.3.10 – Ann Measures the Length of the Rod



Ann measures the length of her rod to be the speed of Bob's sparker multiplied by the coordinate time she measures between the two events. The way Bob measures the length of the rod is similar:

Figure 1.3.11 – Bob Measures the Length of the Rod



We can now use the time dilation formula that relates these two times to determine a relationship between the two measured lengths, *but we have to be very careful here*. Namely, we must ask ourselves, whose measures the dilated time here? Put another way, which of these two observers measures the proper time between the two events? The answer is clearly Bob, since both events occur at the end of his sparker, which is at rest in his frame, meaning that both events occur at the same place – the exact criterion for proper time. Therefore we find that it is Ann's time between events that is longer than the time measured by Bob, giving:

$$\Delta t=\gamma_v\Delta t' \quad \Rightarrow \quad L'=v\Delta t'=v\frac{\Delta t}{\gamma_v}=\frac{L}{\gamma_v} \tag{1.3.7}$$

So Bob measures the rod to be *shorter* than Ann measures it to be (recall $\gamma_v>1$). This phenomenon is known as *length contraction*.

A few comments about this result:

- The longest possible measurement of length occurs in the rest frame of the object whose length is being measured. This length is often referred to as the *proper length*.
- We are accustomed to attributing different visual observations of the same object to optical illusions. This is *not* one of those cases. The same rod *is shorter* for Bob than it is for Ann. Length is not an intrinsic property – it is observer-dependent.
- If Bob zooms by Ann with an identical rod, then Bob will measure Ann's rod to be shorter than his own, *and* Ann will measure Bob's rod to be shorter than hers. We will explore this seeming paradox soon, but the short answer is that *length is not a quality that is inherent to an object*, so the fact that the rods are identical does not mean that their lengths are. Most people are not bothered by the fact that two identical rods may have different colors (due to red/blue shift), because it isn't too difficult to accept that color is not a

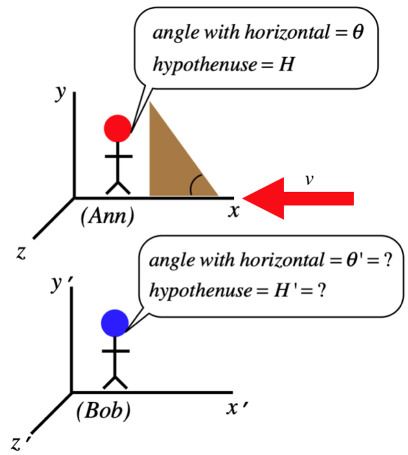
property inherent to objects, but length is significantly tougher to swallow. If it helps, it's okay to say that identical objects have equal proper lengths.

- If we consider examining the same rod in both frames when it is aligned along the y-axis (while relative motion is still along the x-axis), we will fairly easily find that both observers agree on the length. Put another way, *lengths contract only along the dimension parallel to the direction of relative motion*.
- In terms of the lattice points for the two frames, we would say that Bob sees Ann's lattice points ate closer together along the x-direction that his own, and Ann would conclude the same about Bob's lattice points. This is because each can measure the rod in terms of these points, so if the length of the rod is relative, so is the entire space of the moving item.

Angles

Consider our usual setup with Ann and Bob in relative motion along their common x-axis. We know that there is a contraction of length along the x-axis when an object is moving relative to the frame it is viewed in. Furthermore, we know that lengths along the y and z axes do not contract. Suppose Ann has a right-triangular wooden block at rest in her frame as in the figure below, and measures the angle it makes with the x-axis to be θ. What angle does Bob measure?

Figure 1.3.12 – Ann and Bob Measure an Angle



With only the side of the triangle along the x-axis contracting, the angle must change. If we call the length of the base of the triangle in Ann's frame x and the height y, then we get:

$$\theta' = \tan^{-1}\left(\frac{y'}{x'}\right) = \tan^{-1}\left(\frac{y}{\frac{x}{\gamma_v}}\right) = \tan^{-1}\left(\gamma_v \frac{y}{x}\right) = \tan^{-1}(\gamma_v \tan\theta) \tag{1.3.8}$$

Example 1.3.1

Compute Bob's unanswered question in the figure above – what does he measure for the hypotenuse of the triangle, in terms of the hypotenuse and angle measured by Ann?

Solution

Start with the Pythagorean theorem:

$$H' = \sqrt{x'^2 + y'^2} = \sqrt{\frac{x^2}{\gamma_v^2} + y^2} = \sqrt{\left(1 - \frac{v^2}{c^2}\right) x^2 + y^2} = \sqrt{x^2 + y^2 - \frac{v^2}{c^2} x^2}$$

Now write Ann's values of x and y in terms of H and θ:

$$\left.\begin{array}{l} H^2 = x^2 + y^2 \\ x = H\cos\theta \end{array}\right\} \quad H' = \sqrt{H^2 - \frac{v^2}{c^2} H^2 \cos^2\theta} = H\sqrt{1 - \frac{v^2}{c^2}\cos^2\theta}$$

- **2.2: The Nature of Time** by Tom Weideman is licensed CC BY-SA 4.0. Original source: native.

1.4: Paradoxes

The Ladder & Barn Paradox

We have derived some rather outlandish results from the postulate(s) of relativity, using the simple but powerful tool of thought experiments. We will now use that same tool to construct some scenarios that appear at first blush to result in logical inconsistencies in the theory called paradoxes, and then we will shoot them down, demonstrating that the theory is in fact logical and consistent.

The first paradox we will examine involves length contraction, and it goes like this...

A farmer wishes to store a long ladder that he owns inside his barn, but is frustrated to discover that his ladder is too long to fit. Specifically, he finds that is ladder is 50ft long, while his barn is only 40ft long. But like every good farmer, this fellow is well-versed in special relativity, and decides that if he can get the ladder moving fast enough relative to the barn, then the ladder's length will contract enough so that both the front and back barn doors can be closed at the same time while the ladder is inside. Before he actually tries to close the ladder in, he does a test run, zooming the ladder in one door and out the other. Sure enough, as his wife drives their souped-up tractor at 0.6c, he notes that both ends of the length-contracted ladder are briefly within the confines of the two doors:

$$\text{farmer measures ladder length to be: } L' = \frac{L}{\gamma_v} = \sqrt{1-\left(\frac{0.6c}{c}\right)^2}\,(50ft) = 40ft \tag{1.4.1}$$

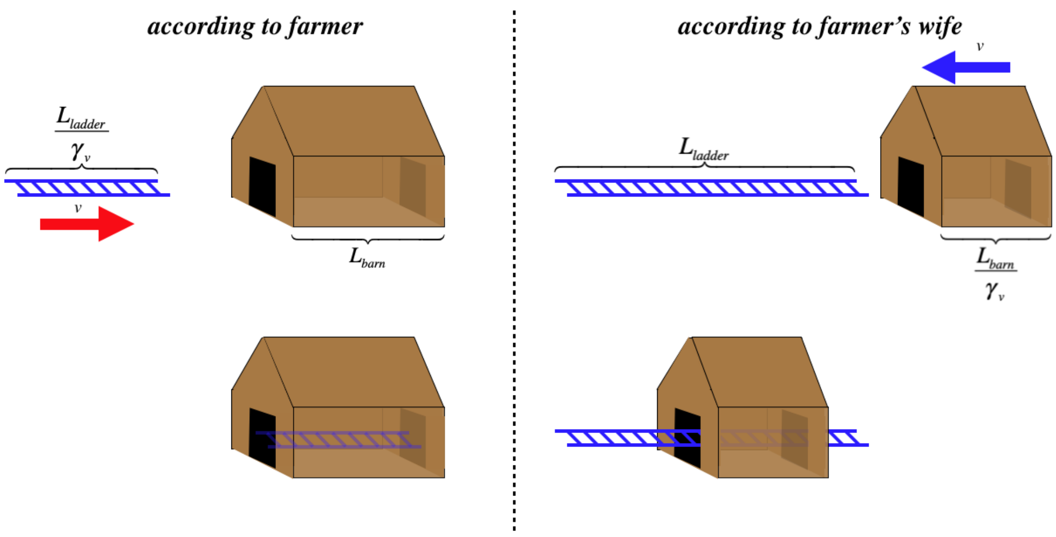
Since the $40ft$ *barn is at rest relative to the farmer, the* $40ft$ *length-contracted ladder just barely fits.*

Overjoyed that his ploy is going to work, he goes to embrace his wife, who looks distraught. When he asks her what is the matter, she says, "When I am zooming into the barn, it is way too short to fit the ladder." Indeed:

$$\text{farmer's wife measures barn length to be: } L' = \frac{L}{\gamma_v} = \sqrt{1-\left(\frac{0.6c}{c}\right)^2}\,(40ft) = 32ft \tag{1.4.2}$$

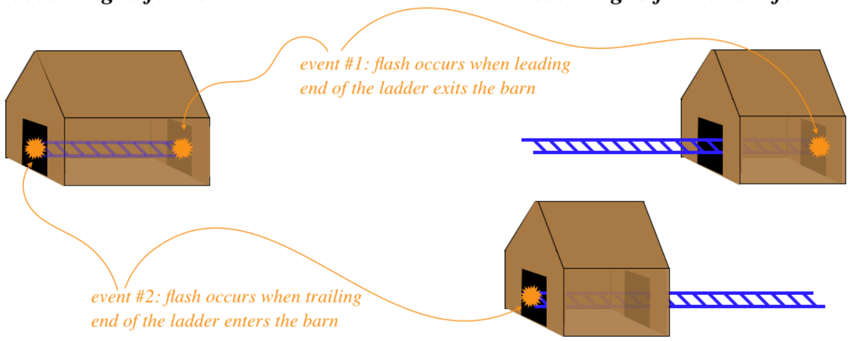
The frame of the farmer's wife sees the ladder at rest and the barn moving past, so naturally the $50ft$ *ladder doesn't fit inside the* $32ft$ *length-contracted barn. So how it is possible that both the farmer and his wife are correct at the same time? How can the ladder both be contained in the barn and not contained at the same time?*

Figure 1.4.1 – Ladder & Barn Paradox



As with everything else in relativity, we can't trust our logic without converting everything into the language of spacetime events. In terms of events, what does it mean for the ladder to be "entirely within the barn?" Well, imagine that each end of the ladder is equipped with a flashbulb that flashes whenever it is in a doorway of the barn. We can say that the ladder is just barely "completely enclosed" if the light at the front of the ladder flashes in the exit doorway at the same instant that the light at the rear of the ladder flashes in the entrance doorway. Then we can say that both ends of the ladder are inside the barn at the same time. But we know from our discussion of simultaneity that what one observer sees as simultaneous events will not be seen as simultaneous by another. So the farmer sees the ladder as being within the barn because he can declare both ends to be within the confines of the barn *at the same time*, even as his wife claims that the front of the ladder exits the barn well in advance of the rear of the ladder entering it.

Figure 1.4.2 – Paradox Resolved By Discarding Simultaneity



Once we accept that the idea of simultaneity is not universal, we realize that "being inside the barn" is a relative concept – two observers don't need to agree on this.

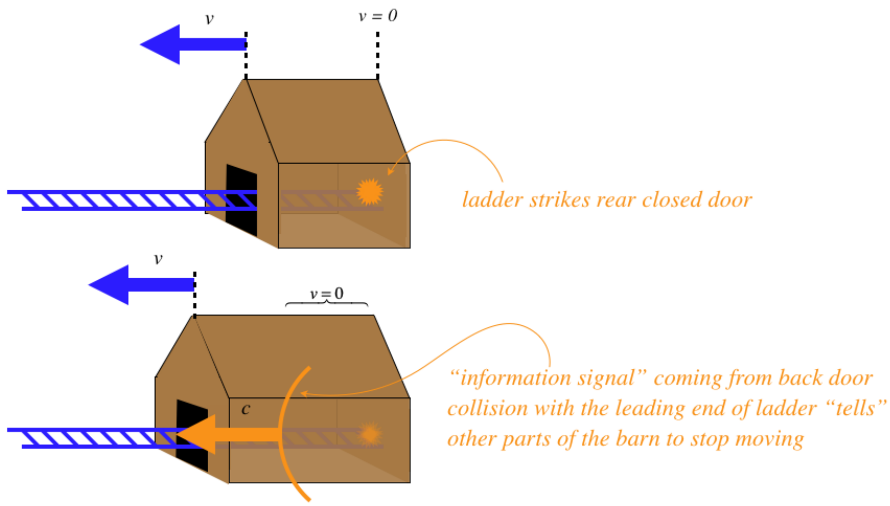
Okay, so much for the “dry run” of driving the ladder into one door and out the other. If the farmer can correctly claim that the ladder is within the barn, what happens if he closes the two barn doors simultaneously, while the ladder is in there? What will his wife see then?

If the barn door closings are simultaneous to the farmer, then they are not so for the wife. She would see the exiting barn door closed first, just as the front of the ladder gets there, and at that moment in her frame, the rear of the ladder will not be in the barn, and the entrance door is not yet being closed. But the question is, if the front of the ladder is stopped by the closed door, how does the rear of the ladder ever get in?

The answer to this is subtle and a bit beyond what we have discussed so far. Before we go into it, it is necessary to point out that everything we have been discussing to now points to a “cosmic speed limit” that is the speed of light. If two frames could move at the speed of light relative to each other, then one would witness time in the other frame to be stopped, and lengths to be shrunk to zero parallel to the direction of motion. This is an asymptotic speed limit that applies not only to solid objects, but also *influences* (such as forces and fields) and *information*.

Given that influences can’t travel faster than c, we must concede that even if we could somehow instantly stop the front of the ladder with the closed barn door, the *rear of the ladder will keep moving forward* until it “gets the message from the front” to stop. From the wife’s perspective, this means that the front of the barn will continue moving at v until the rear of the barn sends a message to it to stop moving. As time passes after the collision of the back door of the barn with the ladder, more and more of the barn will stop moving, until finally the front of the barn gets the message and the entire barn is at rest. If the message to stop could get to the front of the barn instantly, then of course the front door could never be closed and we would have a paradox, since the farmer clearly sees that it is possible to close both doors with the ladder within. So we need to calculate to see if the message is slow enough (even at the cosmic speed limit) to stop the front of the barn before the ladder can be enclosed.

Figure 1.4.3 – Can Both Barn Doors Be Closed?



After t seconds in the wife's reference frame, the barn has gotten longer by an amount $v\Delta t$:

$$\text{length of barn after } t = 32ft + vt = 32ft + (0.6c)\,t \tag{1.4.3}$$

The distance that the light has traveled from the rear door in the wife's reference frame is equal to ct, so the signal reaches the front of the barn when:

$$ct = 32ft + (0.6c)\,t \quad \Rightarrow \quad t = \frac{32ft}{0.4c} \tag{1.4.4}$$

Okay, so we know how long it takes according to the wife for the barn to stop moving, so we can plug this time in to determine the maximum length of the barn:

$$\text{length of stretched barn} = \text{distance the light travels in the total time } = ct = \frac{32ft}{0.4} = 80ft \tag{1.4.5}$$

The ladder is 50ft long in the wife's frame, so it is inside the barn before the front of the barn gets the message to stop moving, and the barn door can be closed, averting the paradox entirely.

The Twin Paradox

The next paradox takes on time dilation. The idea is that if Ann and Bob can both see the other's clock ticking more slowly, what happens if, after awhile, we just bring both of the clocks together (into the same frame) and compare them? Which clock will have elapsed more time?

Two identical twins are born on a space station, and live there for their whole lives, until one of them gets into a spaceship for a long journey. When she speeds away from the space station at a sizable fraction of the speed of light, the twin on the space station (Ann) uses a powerful telescope to look through the window of her sister Bea's spaceship, and sees the effects of time dilation as the clock on Bea's ship is turning much slower than an identical clock in the space station. Bea also has a powerful telescope, and looking back at her sister, see that the space station clock is turning slow compared to her own. When Bea returns from her long journey and reunites with her sister, it would seem that both twins will encounter someone that is much younger than they are, but of course this is impossible. Which one will be older, or will they be the same age at the reunion?

The apparent paradox arises here because of symmetry. Each twin can claim that the other is moving because there is no such thing as a universal inertial frame. What is more, if we define the spacetime events as being the flash of light at the space dock that occurs during both the departure and the return, we see that both twins measure the proper time between these events, because the positions of both events in space are the same.

But *is* this symmetric? Both twins start in the same inertial frame, but in order to end up in different inertial frames, one of them has to *change frames*. For this to happen, she needs to accelerate, and this (at least while she is accelerating) puts her into a special non-inertial frame, which she *can* do an experiment to detect. The flash of light at the space dock occurs at the same position in each twin's frame, so they both measure a proper time, but only Ann measures the spacetime interval between those flashes, as only she stays in the inertial frame of the clock that records both events.

This means that the viewpoint of Ann remains correct – her sister experiences fewer years than she does during the trip. They started as twins, but after the trip Ann is several years older.

Alert

It is important to understand that this is not the fountain of youth – both twins age in the usual way. It isn't the rate of aging that is changed, it is the rate of time flow itself that is different. The older sister has more experiences than the younger one, and if they both live to be 100, they have equally-long lives. Their lives just don't conclude at the same time, even though they started at the same time.

Let's do a specific calculation of this effect. Suppose that before the trip begins, Bea tells Ann that she's going to a destination that is 20 light-years away and coming back. Her ship will travel at a more-or-less constant speed of $0.8c$ (except for the quick accelerations near the two endpoints). Ann does the simple math and determines that to cover the distance of the round trip at that speed will require 50 years:

$$\Delta x = v\Delta t \quad \Rightarrow \quad \Delta t = \frac{\Delta x}{v} = \frac{40 \; light \; years}{0.8c} = 50 \; years \tag{1.4.6}$$

And this is in fact the amount of time that Ann measures for Bea's trip. But what does Bea measure? Once Bea quickly gets up to speed, the distance isn't so far anymore, as the separation of the two endpoints is length-contracted (think of placing a very long rod between the endpoints). What is 20 light-years away in Ann's frame in Bea's frame is:

$$L' = \frac{20 \; light \; years}{\gamma_v} = \sqrt{1-\left(\frac{0.8c}{c}\right)^2} \left(20 \; light \; years\right) = 12 \; light \; years \tag{1.4.7}$$

At a speed of $0.8c$, this 24 light-year round trip only takes 30 years. So Bea returns home to find that her twin sister is 20 years older than she is.

Let's call the departure of Bea and her arrival spacetime events A and B. Note that these both occur at the same place in space (Ann & Bea's home), which means that *both* Ann and Bea measure a proper time between the events. The fact that these are different comes back to our discussion around Equation 1.2.9 – the proper time measurement depends upon the history of the clock that measures it. Ann remains in an inertial frame, but Bea must accelerate to leave, turn around, and stop when arriving home. So the integrals come out different for the two sisters. This appeal to pure mathematics is perhaps not very satisfying, so let's try a new thought experiment to try to clear it up. This involves three players: Ann, Bob, and Chu, who have the following roles:

- Ann will be our "reference" observer. We will watch the motions of the other two through a telescope from the comfort of her (by her account, stationary) inertial frame on her spaceship.
- Bob will be engaging in a race from the starting point to the finish in a spaceship that travels in a straight line at a constant rate (as measured by Ann) of $0.5c$. So Bob remains in an inertial frame throughout the trip.
- Chu will be the other contestant in the race, but he will get off to a slow start with a speed of $0.25c$. Sometime during the race he will instantaneously speed his spaceship up to $0.75c$ (both speeds according to Ann) in an effort to catch up to Bob before the finish line.

All three participants have identical clocks on board their ships, and when the race begins, a flash of light (spacetime event) at the start line signifies the start of the race. When a ship crosses the finish line, another flash of light is given off there, providing a spacetime event that indicates that the ship has finished the race. Ann measures the distance between the start and finish lines as being 40 light-minutes (roughly the distance from our Sun to Jupiter).

Let's compute the time it takes Bob to complete the race, according to Ann. This is easy - no relativity necessary. Traveling at $0.5c$ over a distance of 40 light-minutes requires 80 minutes.

According to Ann, Chu makes his shift in speeds 40 minutes into the race (when Bob is halfway to the destination). So in the first 40 minutes at a speed of $0.25c$, Chu travels 10 light-minutes. Then for the remaining 30 light-minutes of the trip he is traveling at $0.75c$, which means it takes him 40 more minutes... He finishes in a tie with Bob!

The start and finish of the race both occur at the same position in Bob's frame (the front tip of his ship), and he is in an inertial frame, so he measures the proper time interval $\gamma_{0.5c}\Delta\tau_{Bob} = \Delta t_{Ann}$, as we know from the time dilation formula between two inertial frames. Bob's time comes out to be:

$$\Delta\tau_{Bob} = \frac{\Delta t_{Ann}}{\gamma_v} = \frac{\sqrt{3}}{2}(80min) \approx 69min \tag{1.4.8}$$

Now let's compute the time measured by Chu. We can do this by splitting his trip into three events. We already have two of them – the flash that occurs at the start and finish. As with Bob, these two events occur at the same place in his frame, so he measures a proper time, but as his inertial frame changes, it is not the same. The third event we will define as a flash at the front tip of Chu's spaceship when he suddenly changes speeds. Now we can compute his time for each leg of the trip separately. He remains in an inertial frame during each leg, so we can use the time dilation of Ann for each leg (like we did for Bob's whole trip):

$$\begin{aligned}\Delta\tau_{Chu} = \Delta\tau_{Chu}\,(leg\ 1) + \Delta\tau_{Chu}\,(leg\ 2) = \frac{\Delta t_{Ann}\,(leg\ 1)}{\gamma_{v_1}} + \frac{\Delta t_{Ann}\,(leg\ 2)}{\gamma_{v_2}} = \sqrt{1-0.25^2}\,(40min) \\ + \sqrt{1-0.75^2}\,(40min) = \frac{\sqrt{15}+\sqrt{7}}{4}(40min) \approx 65min\end{aligned} \tag{1.4.9}$$

So we see that less time elapses for Chu than for Bob during this race. So how does this apply to the twin paradox? Let's view the whole race from Bob's perspective. Let's call the direction of the race the $+x$-direction. According to Bob, he and Chu start at the same position, and Chu is initially going in the $-x$-direction (because Bob is moving faster than Chu in the $+x$ direction according to Ann). A little while later, Bob sees Chu suddenly stop and immediately start coming back toward him (because Chu is now moving faster than Bob in the $+x$ direction according to Ann), until they are back at the same position. Bob insists he was stationary the whole time, so according to Bob, Chu basically took off and came back. And sure enough, when he came mack, Chu had aged 4 minutes less than Bob.

While we haven't proven it with this one example, we do see a general result: For the same two spacetime events, all proper time measurements (like Bob's and Chu's) are shorter than all coordinate time measurements that are not proper (like Ann's). Also, the proper time interval between two events measured in an inertial frame (like Bob's) is *longer* than all of the other (non-inertial) proper time measurements (like Chu's).

CHAPTER OVERVIEW

2: Kinematics and Dynamics

2.1: Spacetime Diagrams

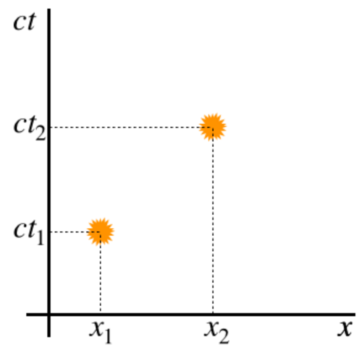
Graphing Spacetime Events

Up to now, we have been representing the position of a spacetime event relative to the spatial axes of Ann and Bob, while representing time by drawing the axes at different positions, with the set of axes remaining fixed being the coordinate system of the observer from whose perspective we are seeing the event. Here we will represent spacetime events in a more efficient and useful (though perhaps a bit more abstract) manner.

We will stay with simplified situations where all the action occurs along the x-axis, which means that we have room for something else. Specifically, we will draw the position (along x-axis), and *time* axes for an observer in an inertial frame whose perspective we are viewing from. With this construction, a spacetime event is simply a point located in the plane of the two axes, and the coordinate position and coordinate time for the observer we are viewing from can be read off the axes directly.

In 9HA we worked with 1-dimensional graphs with the position represented on the vertical axis, and the time on the horizontal axis. In relativity, it is standard (perhaps just to alert the reader to the fact that relativity is being considered) to use time as the vertical axis. What is more, these diagrams give both axes the same units by scaling the vertical axis by the speed of light, c. The resulting representation of spacetime events is called a *spacetime or Minkowski diagram*.

Figure 2.1.1 – Spacetime Events on a Minkowski Diagram



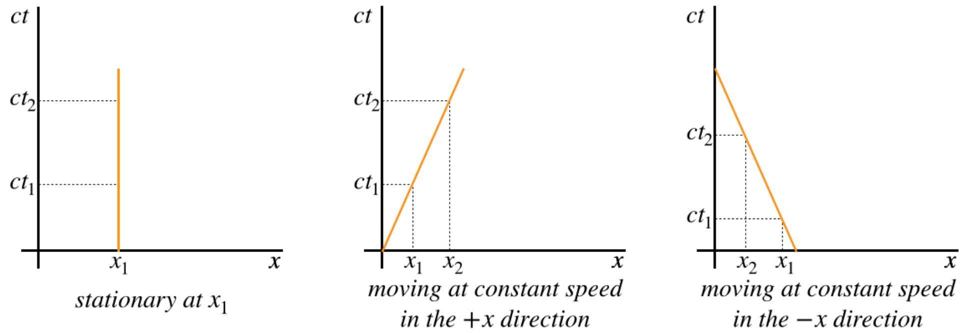
The times indicated by the values on the vertical axis are those measured by synchronized clocks in the frame of the observer represented by these axes. If the two events happen to be aligned vertically, then the two events occur at the same position in this frame, which means the difference $t_2 - t_1$ in that case is also the proper time interval $\Delta\tau$. The spacetime interval is related to the "distance" separating these events in the diagram. If the spacetime interval is measured in an inertial frame, then it is the "length" of the straight line connecting the events, otherwise it is the "arclength" of the curve connecting the events. All of these quoted words must be very confusing at the moment, but things will be made clearer below.

World Lines

While it is nice to be able to draw two separate events on the same diagram, rather than drawing multiple drawings as we did before, there is an even greater advantage to the tool of the Minkowski diagram – we can represent *spacetime trajectories*, also known as *world lines*. Suppose we track a flashing beacon as it moves along the x-axis. Each flash records a spacetime event with the position it is located and the time of the flash, so we get a string of several points on the spacetime diagram. If the flashing frequency is increased to infinity, then the spacetime points form a continuous curve that tracks the spacetime history of the moving object.

Let's consider a few special world lines. When someone observes a stationary object, its trajectory in the Minkowski diagram must be such that the position doesn't change (but of course the time does). If this person witnesses an object moving at a constant speed, then the world line is sloped, positively when the motion is in the $+x$ direction, and negatively when the motion is in the $-x$ direction.

Figure 2.1.2 – Some Simple World Lines



A closer look at the values of the slope of a world line reveals that it is:

$$\text{slope of a world line} = \frac{ct_2 - ct_1}{x_2 - x_1} = \frac{c}{u}\ , \quad u \equiv \frac{\Delta x}{\Delta t} = \text{speed of the object measured by observer} \tag{2.1.1}$$

That is, the inverse of the slope of the world line of an object is the fraction of the speed of light that the observer measures for that object, with the sign representing the direction of motion. If the "object" happens to be a light beam, then its slope is ± 1.

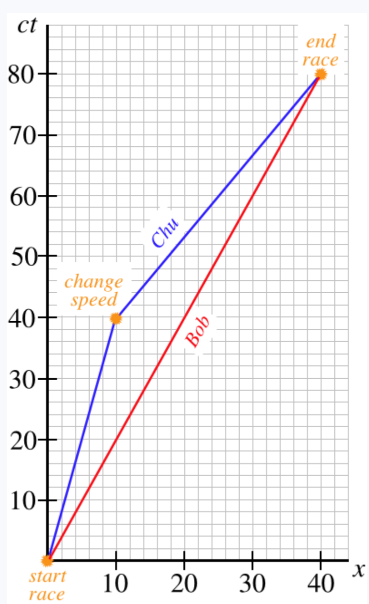
Example 2.1.1

Let's return to the thought experiment for the twin paradox involving Ann, Bob, and Chu at the end of Section 1.4, and use a spacetime diagram to depict the race.

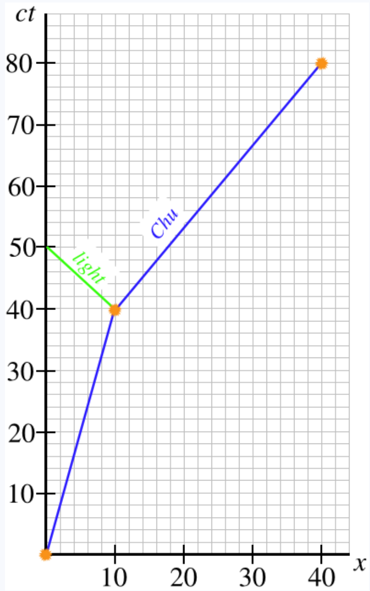
a. *Draw a spacetime diagram in Ann's reference frame depicting the world lines of both Bob and Chu, and label the important spacetime events along these worldlines.*
b. *Use the diagram to determine the time on Ann's clock in her spaceship (not at the lattice point in her reference frame) when she **sees through her telescope** that Chu has changed speed.*
c. *Use the diagram to determine the time on Bob's clock in his spaceship when he **sees through the window of his spaceship** that Chu has changed speed.*

Solution

a. The diagram below has light minutes as units on both axes. Both start the race at the same point in spacetime ($x = 0$, $t = 0$ in Ann's frame), and end at the same point in spacetime ($x = 40$ light-minutes, 80 minutes later). Chu changes speed at $x = 10$ light-minutes, after $t = 40$ minutes. Bob moves at a constant speed of half the speed of light, so the slope of his world line is 2. Chu moves at one-quarter the speed of light and then three-quarters the speed of light, so the slopes of his world line segments are 4 and $\frac{4}{3}$, respectively.

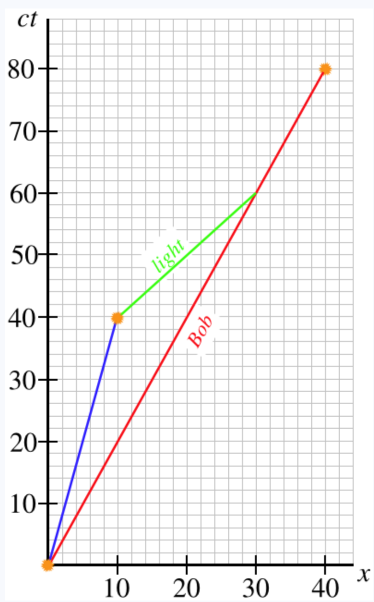


b. Ann doesn't see the event of Chu changing speed until light from that event reaches her. The world line of this light has a slope of -1 because it is coming back to Ann (in the $-x$-direction). Drawing this into the diagram shows that she discovers Chu has changed speeds at time $t=50$ minutes, because that's when the light world line intersects with Ann's world line (which is the vertical axis, since she starts at her origin and never moves).



c. Now we want to know when Bob sees light from the event of Chu changing speed. This time the light goes in the $+x$-direction to get to Bob, so it has a slope of $+1$. This allows us to find the intersection point of the light world line with Bob's world line, but then deriving from that the time on Bob's clock requires a little more work. The simplest way to get this number is to find what the time is on Ann's clock, then use the time dilation formula to get Bob's time. Looking at the diagram below, we see that the time of this intersection in Ann's frame is 60 minutes, which in Bob's frame translates to:

$$t_{Bob}=\frac{t_{Ann}}{\gamma_v}=\sqrt{1-0.5^2}\,(60\ minutes)\approx 52\ minutes$$



Minkowski Spacetime

Let's represent on spacetime diagrams a comparison of what Ann sees to what Bob sees when they witness the same object's motion. Specifically, let's say they both observe the same clock, which gives off two light flashes at different times, and records those times. Let's further have the clock remain stationary in Ann's frame between these flashes.

With the clock at rest in Ann's frame, the world line for it looks like the first graph in the figure above. With Bob moving in the $+x$-direction relative to Ann at a speed v, he sees this same object moving in the $-x$-direction at a speed v, which means the worldline he sketches for the object looks like the third graph in the figure above.

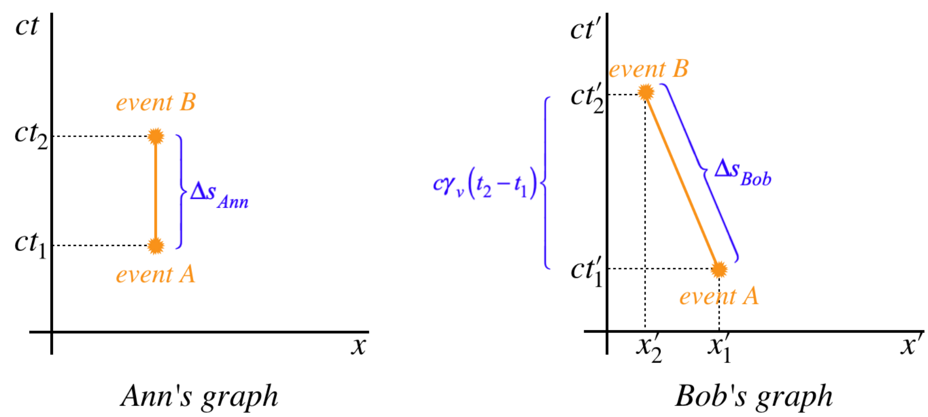
As we showed above, the spacetime interval between the two events is the *square of the length of the segment of the world line* connecting the two events in Ann's spacetime diagram. When we first discussed the nature of time, we made a point of noting that all proper times, and therefore the spacetime interval, are invariants, which means that everyone measures the same value. This would seem to indicate that if Bob measures the length of the segment of the world line he draws between the two events, he should get the same result. In fact this is true, but as we will see, there is a surprise twist.

We found that the time dilation effect gives the comparison between time intervals to be (Bob is the primed frame here):

$$t_2' - t_1' = \gamma_v \left(t_2 - t_1\right) \tag{2.1.2}$$

But wait, since $\gamma_v > 1$, this means that $ct_2' - ct_1' > ct_2 - ct_1$, and when we look at the spacetime diagrams for Ann and Bob, we see this means that the vertical change in the graph is greater for Bob than it is for Ann.

<u>Figure 2.1.3 – Comparison of Spacetime Intervals for Ann and Bob</u>



It sure doesn't look like there is any way that Ann and Bob could agree on the length of the spacetime interval between these two events! Now for the twist (and the explanation for why we used the quotes on the words "distance", "length", and "arclength" earlier). Let's go back to the light clock thought experiment from Section 1.2. The distance traveled by the light beam was, according to Ann and Bob:

$$\begin{aligned} Ann: &\quad c\Delta t = 2L \\ Bob: &\quad c\Delta t' = \sqrt{\left(2L\right)^2 + \left(x_2' - x_1'\right)^2} \end{aligned} \tag{2.1.3}$$

Eliminating the $2L$ from these equations gives:

$$c^2\Delta t^2 = c^2\Delta t'^2 - \Delta x'^2 \tag{2.1.4}$$

The left side of this result is the spacetime interval squared written in Ann's frame, so with the spacetime interval squared being invariant, the right side of the equation must be how it is written in Bob's frame. This looks very close to the Pythagorean theorem, with the exception that the squares that are added are not all positive. Now we can see how the lengths of the world line segments can actually be equal after all. The horizontal leg of the right triangle in Bob's graph actually contributes a *negative* amount to the length of the hypotenuse. Clearly we have redefined "length" in this context (which we no longer call length, preferring the word "interval" to avoid confusion), but it is necessary to do so to accommodate the nature of time and the constancy of the speed of light.

We therefore reiterate explicitly what we said about the invariance of the spacetime interval back in Section 1.2:

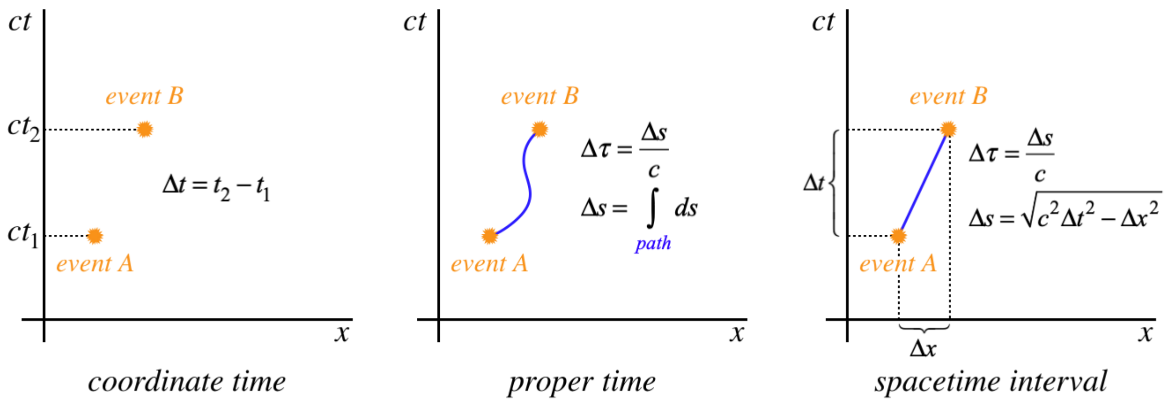
$$\Delta s'^2 = \Delta s^2 \quad \Rightarrow \quad c^2\Delta t'^2 - \left(\Delta x'^2 + \Delta y'^2 + \Delta z'^2\right) = c^2\Delta t^2 - \left(\Delta x^2 + \Delta y^2 + \Delta z^2\right) \tag{2.1.5}$$

Notice this works for the case of Ann and Bob above, since $\Delta x = 0$ for Ann.

Three Measurements of Time Between Two Events

With the tool of spacetime diagrams and a means for computing intervals between events, we can gain a better understanding of the three versions of time.

Figure 2.1.4 – Three Measurements of Time in Spacetime Diagrams



The figure above shows the same two events expressed in the same frame of reference. The first graph shows that the coordinate time elapsed between the two events is simply the difference in the times measured by the synchronized clocks at the two lattice points.

The middle graph expresses a general version of proper time. The time is measured by a clock whose world line passes through both events, and it equals the path length of the clock through spacetime divided by c. This path length is found by doing a line integral to add up all the tiny contributions ds – each of which is an infinitesimal straight line segment – so:

$$\Delta s = \int\limits_{path} ds = \int\limits_{path} \sqrt{c^2dt^2 - dx^2} \tag{2.1.6}$$

The right graph is merely a special case of the middle graph. Integrating ds to get the path length in that case just gives the length of a straight line (in Minkowski space). Note that the velocity of the clock is changing in the middle graph (the slope of its world line is changing), while in the right graph it is constant. Put another way, the clock is not in an inertial frame in the middle graph, while it is in the right graph.

If a different observer were to graph spacetime coordinate points for the same two events, and then watch the same clocks, the three graphs would look different. The coordinate time separating the two events would change. The world lines of the clock would look different in the other two graphs as well. The world line in the right graph would still be a straight line, but with a different slope. The world line would be different in the middle graph as well, but would still be curved. However, the calculation of Δs in these two graphs would come out the same as they did for the original observer.

Alert

The calculation of Δs for the moving clock in the noninertial frame (middle graph) is the same when computed by any observer. The calculation of Δs for the moving clock in the inertial frame (middle graph) is the same when computed by any observer. Don't confuse this for saying they are the same as each other! The time elapsed on the clock is proportional to the path length Δs through spacetime, and the path length is different for a straight line compared to a curved one.

Here we will repeat something we discussed at the end of Section 1.2, as it may make more sense now that we are armed with the visual aid of the spacetime diagram. We can rewrite Equation 2.1.6 by pulling $c\ dt$ out of the radical:

$$\Delta s = \int\limits_{path} \sqrt{1 - \frac{1}{c^2}\frac{dx^2}{dt^2}}\ c\ dt = \int \sqrt{1 - \frac{u(t)^2}{c^2}}\ c\ dt \tag{2.1.7}$$

The function $u(t)$ is the velocity of the clock as a function of time, which in general can vary during the clock's journey through spacetime. If it does vary, then the radical needs to be kept in the integral. [*Naturally "varying velocity" means that such a clock is not remaining in an inertial frame.*] If the velocity remains constant, then the radical can be removed from the integral, giving simply:

$$\Delta s=\sqrt{1-\frac{u^2}{c^2}}\,c\Delta t \quad\Rightarrow\quad \Delta\tau=\frac{\Delta s}{c}=\frac{\Delta t}{\gamma_u} \tag{2.1.8}$$

We recognize this as the time dilation relationship between the proper time $\Delta\tau$ in one inertial frame and the coordinate time Δt in an inertial frame moving at a relative speed of u.

We found in our discussion of the twin paradox that the proper time span between two events is longer when measured using an inertial clock than when measured with a non-inertial clock (the twin in the non-inertial space ship ages less over the round trip). In other words, the value of Δs in the middle graph of Figure 2.1.4 is less than the value of Δs in the right graph (remember, the world lines connect the same two events). This seems weird in light of the old adage, "the shortest distance between two points is a straight line," but remember that this is Minkowski space with the weird version of the Pythagorean theorem. To see clearly that this is true, we can change frames to one were the two events are at the same value of x – remember that these Δs values don't change when we change frames. This will change the right graph such that the world line is a vertical straight line, while the middle graph remains curvy. Because of the $-dx^2$ contributions to the integral, any horizontal deviations from the vertical path will serve to *reduce* the magnitude of the integral. So the proper time $\Delta\tau=\frac{\Delta s}{c}$ for a straight world line between two events (i.e. measured by an inertial clock) is greater than the $\Delta\tau$ for any non-straight world line between the same two events (i.e. measured by a non-inertial clock).

Example 2.1.2

Once again returning to the thought experiment for the twin paradox involving Ann, Bob, and Chu at the end of Section 1.4, use the spacetime diagram created in part (a) of Example 2.1.1 above to answer the following:

a. *Compute the (Minkowski) length of the world line for Bob to get the time elapsed for him during the race.*
b. *Repeat the process in part (a) for Chu.*

Solution

a. The x and ct parts of Bob's world line are 40 and 80 light-minutes (lm), respectively, so the Minkowski length of his world line is:

$$\Delta s=\sqrt{80^2-40^2}\ lm\approx 69\ lm$$

Dividing Δs by the speed of light gives the time, which means that 69 minutes elapse for Bob. This agrees with the result we got in Section 1.4.

b. To compute the length of the world line for Chu, we need to compute the lengths of two separate segments and then add them together.

$$\left.\begin{array}{lll}\Delta x_1=10\ lm\ ,\ \ c\Delta t_1=40\ lm & \Rightarrow & \Delta s_1=\sqrt{40^2-10^2}\ lm\approx 38.7\ lm\\ \Delta x_2=30\ lm\ ,\ \ c\Delta t_2=40\ lm & \Rightarrow & \Delta s_2=\sqrt{40^2-30^2}\ lm\approx 26.5\ lm\end{array}\right\}\quad \Delta\tau_{Chu}=\frac{\Delta s_1+\Delta s_2}{c}\approx 65min$$

Again, this is in agreement with our result before.

Causality

You may have noticed that there is a peculiarity lurking in what we have discussed in this section. In case you haven't, consider the following. We start with events at two points in spacetime, and we want to measure the spacetime interval time that elapses between these events. So we start a clock at one of the events, and move it at a constant speed to the other, recording the times of each event. Combining Equation 2.1.1 with Equation 2.1.6, we see that the clock must record this time span (omitting the Δy^2 and Δz^2 terms):

$$\Delta\tau=\frac{\Delta s}{c}=\frac{\sqrt{c^2\Delta t^2-\Delta x^2}}{c} \tag{2.1.9}$$

Naturally these two events can occur anywhere in spacetime, so the question is this: If the position in spacetime is such that $\Delta x^2 > c^2\Delta t^2$, exactly what time does the clock measure?! Does an imaginary 'i' suddenly appear in the digital readout?

When we say that a clock cannot be moved from one event to another to record the proper time between the events, what we are really saying is that *there is no world-line that connects them*. When we think about a spacetime diagram, this means that we can place individual events wherever we like, but we can only connect them with world lines if such a line has a slope whose absolute value is greater than 1 (a speed smaller than c). This leads to three different categories for how two spacetime events relate to each other:

time-like separation: $\Delta s^2 > 0$

This is the case we have been talking about before now, where an object can follow a world line and connect the two events. There is a well-defined proper time associated with the spacetime interval. When two events are separated in this way, we can always find an inertial frame where the events occur at the same spatial position.

light-like separation: $\Delta s^2 = 0$

In this case, the world line that connects the two events must belong to a light beam. It isn't possible to change the reference frame to one that is moving at the speed of light relative to another, so although there is a world line that can connect them, it can't belong to a clock.

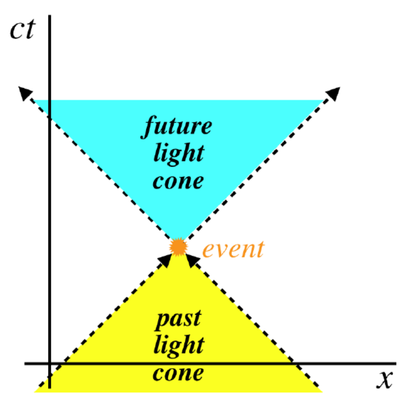
space-like separation: $\Delta s^2 < 0$

For this case, not even light can connect the two events with a world line. If we imagine an event occurring because someone at the spatial position of the event pushes a button to cause a light to flash, then the people responsible for two events separated in this way have no idea that the other event exists, as there is no world line to carry a message about one event to the other event.

This last point about space-like separated events brings up an important concept – that of *causality*. The study of physics is all about cause-and-effect – it is a net force that *causes* an acceleration, for example. We see now that we can relate two events according to whether one can cause the other. If there is no way, even in principle, for a message to get from one event to another, then there is no way for the first event to cause the other.

We can characterize what events can be causally-related nicely by using a spacetime diagram. Start by choosing a spacetime event. Then note that to be causally-related to this event, a second event must be positioned such that a straight world line drawn between the two has a slope that is no less than 1 (and if it has a slope of 1, they must be related through a signal that moves at the speed of light). This limits the region of the second event relative to the first into what is called the *light cone* of that first event.

Figure 2.1.5 – Light Cone of an Event



The light cone divides into two halves, called the *future* and *past light cones*. Only events in the past light cone can be a cause for the event at the vertex of the light cone, and the event at the vertex of the light cone can only be a cause for events within the future light cone.

It is important to keep in mind that a spacetime diagram is *based on a specific frame of reference*, and if we change to another frame, the relative positions of the events change (see Figure 2.1.3, for example). But because the spacetime interval is an invariant, what lies within the future(past) light cone of an event in one inertial frame must lie within the future(past) light cone in all inertial frames. We can put this in a far less abstract way: If in Ann's frame event #1 occurs before event #2, and the two events lie within each other's light cones, then event #1 also occurs before event #2 in Bob's frame, no matter what the relative speed of Ann and

Bob may be. This assures that if Ann declares that event #1 has caused event #2, there's no observer that can disagree with this on the grounds that event #2 has occurred first.

For a pair of events that do not lie within a light cone (such as two events that are simultaneous in one frame but at different positions, so that they are horizontally-aligned in the spacetime diagram), then the ordering of events is not necessarily preserved in all frames. To see this, let's return to the ladder-and-barn paradox from Section 1.4.

Suppose the farmer's wife is driving one tractor in the $+x$ direction while the farmer's daughter is driving another tractor in the $-x$ direction in an effort to get two ladders into the barn at once (both barn doors are open, and they are coming in from opposite directions. According to the farmer, both ladders can be inside the barn at the same time, because the events at the barn doors are simultaneous in his frame. [They are at different positions, so these two events lie outside the light cone from each other.] The farmer's wife notes that the front of her ladder reaches the back of the barn before the rear of her ladder enters, so she says that the event at the back door occurs before the event at the front door. The farmer's daughter sees the exact opposite – the event at the front door occurs first for her.

There is no way that the farmer's wife can ever claim that the event at the back door *causes* the event at the front door (even though it comes first), because her daughter would argue that it is impossible for the event that occurs second to cause the event that occurs first. Recall that the fact that a light signal cannot get from the back door event to the front door event is exactly what resolved the paradox of whether both doors can be closed with the ladder inside.

2.2: Lorentz Transformation

Transformations Between Inertial Frames

When we first studied relative motion in Physics 9HA, we wrote down a way of translating between the values measured in the two frames. This set of equations was called the Galilean transformation equations. As sensible as these are, they clearly are not correct in light of what we now know about relativity. Most notably, the Galilean transformation assumes a universal time variable that is common to all frames.

So now we seek a new set of transformation equations to relate the spacetime coordinates of frames in relative motion. We will start with a couple of simplifying assumptions. First, the two frames in question share a spatial origin at the moment in time we will call $t=t'=0$ – we will define "event A" as occurring at this spacetime point. The effect of doing this is that distances and time intervals between this event and a second event are now just the spacetime coordinates themselves. For example: $\Delta x=x-0=x$ and $\Delta t'=t'-0=t'$.

For our second assumption, we will continue to define the relative motion as the primed frame moving at a speed v in the $+x$-direction relative to the unprimed frame.

In order to get a set of equations that gives us a translation between the $(ct,\ x,\ y,\ z,)$ spacetime coordinates measured in one frame and the $(ct',\ x',\ y',\ z',)$ spacetime coordinates measured in the other, we begin by noting that with motion only along the x-axis, the y and z coordinates will remain unchanged. For example, we know that lengths along those directions do not contract, so we would not expect the coordinates to be related in any way other than $y'=y$ and $z'=z$. But what about the x and ct coordinates?

We start by assuming that the transformation is a *linear* one, not unlike the Galilean transformation (after all, the Galilean transformation *does* work for frames whose relative speed is low). This means that the primed values can be written as linear combinations of the unprimed values:

$$\begin{aligned} x' &= J\cdot x + K\cdot ct \\ ct' &= L\cdot x + M\cdot ct \end{aligned} \tag{2.2.1}$$

Our goal is to determine the unknown constants J, K, L, and M that work for relativity. Let's start by defining "event B" viewed by the primed observer. Let's say that this event occurs at this observer's time t', and takes place at the origin of the unprimed frame. Since the primed observer sees this frame moving in the $-x'$-direction for a time period of t' after starting at the origin, the primed observer sees this event occur at the position $x'=-vt'$. Plugging $x=0$ (the event occurs at the unprimed origin) into the first equation above and comparing gives us the constant K:

$$\left.\begin{aligned} x' &= -vt' \\ x' &= 0 + K\cdot ct \end{aligned}\right\} \;\; K = -\left(\frac{v}{c}\right)\left(\frac{t'}{t}\right) \tag{2.2.2}$$

Events A and B both occur at the origin of the unprimed frame, so the time span between them is the proper time, and the frame is inertial, so it is the spacetime interval. Therefore the time measured between these events in the primed and unprimed frames are related according to the usual time dilation formula:

$$t'=\gamma_v t \tag{2.2.3}$$

Plugging this in above gives us the constant K:

$$K=-\left(\frac{v}{c}\right)\gamma_v \tag{2.2.4}$$

Using this same event B, we can obtain the constant M as well. Plugging in $x=0$ gives:

$$ct'=0+M\cdot ct \quad\Rightarrow\quad M=\frac{t'}{t}=\gamma_v \tag{2.2.5}$$

Recapping what we have so far:

$$\begin{aligned} x' &= J\cdot x - \left(\frac{v}{c}\right)\gamma_v ct \\ ct' &= L\cdot x + \gamma_v ct \end{aligned} \tag{2.2.6}$$

Now to determine the other two constants, define "event B" as occurring at the origin of the primed frame, $x'=0$. The unprimed observer will see this event occur at the position $x=vt$, which we can plug back in to get:

$$0=J\cdot vt-\gamma_v vt \quad \Rightarrow \quad J=\gamma_v \tag{2.2.7}$$

To find the final constant L requires noting that the time measured in the primed frame for event B is now the proper time, and a bit more algebra than was needed for the previous constants (which is omitted here):

$$\left.\begin{array}{l} ct'=L\cdot vt+\gamma_v ct \\ t=\gamma_v t' \end{array}\right\} \; L=-\left(\frac{v}{c}\right)\gamma_v \tag{2.2.8}$$

Putting everything together gives us the *Lorentz transformation equations*:

$$\begin{aligned} ct' &= \gamma_v\left[ct-\left(\frac{v}{c}\right)x\right] \\ x' &= \gamma_v\left[x-\left(\frac{v}{c}\right)ct\right] \\ y' &= y \\ z' &= z \end{aligned} \tag{2.2.9}$$

The symmetry between the x and t variable is apparent, and shows the important difference between relativity and galilean physics – time is not universal and unaffected by the position of an event. Notice that when the velocity is very small compared to the speed of light (as it is in our everyday experience), then letting $\frac{v}{c}\rightarrow 0$ changes the Lorentz transformation equations into the Galilean transformation equations.

Finally, it should be noted that these transformations can also be written in terms of *changes* in these variables from one event to another. In effect, this is hidden in the equations themselves, as event A simply has all the variables equal to zero.

These equations give the spacetime coordinates of an event in the primed frame given the spacetime coordinates of the same event in the unprimed frame. But what if we want to do the reverse – find the coordinates of the event in the unprimed frame from those in the primed frame? [*This is called the inverse of this transformation.*] It's actually quite easy to do – the only difference in perspectives between these two frames is the sign of the velocity. We get the inverse transformation by simply replacing the v everywhere in the equations with $-v$.

Example 2.2.1

We have said that the interval-squared $\Delta s^2=c^2\Delta t^2-\Delta x^2-\Delta y^2-\Delta z^2$ is an invariant, which means that it is the same in every inertial frame. Use the Lorentz transformation equations to show that this is true.

Solution

We want to show that $\Delta s'^2=\Delta s^2$, which makes this a pure plug-in. Clearly the y and z changes are equal in both frames, so we will ignore them and deal with just the t and x changes:

$$\begin{aligned} \Delta s'^2 &= c^2\Delta t'^2-\Delta x'^2 \\ &= \left(\gamma_v\left[c\Delta t-\left(\frac{v}{c}\right)\Delta x\right]\right)^2-\left(\gamma_v\left[\Delta x-\left(\frac{v}{c}\right)c\Delta t\right]\right)^2 \\ &= \gamma_v^2\left[\left(c^2\Delta t^2-\cancel{2v\Delta x\Delta t}+\frac{v^2}{c^2}\Delta x^2\right)-\left(\Delta x^2+\cancel{2v\Delta x\Delta t}+v^2\Delta t^2\right)\right] \\ &= \cancel{\gamma_v^2}\left[\cancel{\left(1-\frac{v^2}{c^2}\right)}\left(c^2\Delta t^2-\Delta x^2\right)\right] \\ &= \Delta s^2 \end{aligned}$$

Revisiting Previous Results

After all that struggle with thought experiments and spacetime diagrams, only now do we have a simple, powerful tool for achieving the same results. Time dilation is downright trivial. If (unprimed) Ann sees two events occur at the same place ($\Delta x=0$) separated by a time interval Δt, then the time span that (primed) Bob measures between these events is:

$$c\Delta t' = \gamma_v \left[c\Delta t - \left(\frac{v}{c}\right) \cancelto{0}{\Delta x}\right] \quad \Rightarrow \quad \Delta t' = \gamma_v \Delta t \tag{2.2.10}$$

We can also look at simultaneity. Events that are simultaneous in Ann's frame ($\Delta t = 0$) are not simultaneous in Bob's:

$$c\Delta t' = \gamma_v \left[c \cancelto{0}{\Delta t} - \left(\frac{v}{c}\right) \Delta x\right] \neq 0 \tag{2.2.11}$$

Looking at this expression, we also see that $\Delta t'$ is negative (i.e. $t_2 < t_1$) when $\Delta x'$ is positive (i.e. $x_2 > x_1$). This means that for the two events that Ann sees as simultaneous, Bob sees the event with the greater x-value as occurring first (note that we are still assuming that Bob is moving in the $+x$-direction relative to Ann). So is Ann flies by Bob in a spaceship where she sees lights on the front and rear of her ship flashing in sync, Bob sees the light on the rear of her ship flashing ahead of the light on the front.

Reproducing length contraction is a bit more difficult to obtain from the Lorentz transformation equations. the reason is that the length that is measured by one observer depends upon different events than the length measured by the other observer. That is, the length of an object in a given frame is the distance between events located at both ends of the object *that occur at the same time*, and as just noted, events simultaneous in one frame are not simultaneous in the other. Nevertheless, we can get the length contraction result with some care.

Two events that are simultaneous at both ends of an object according to Bob gives:

$$0 = c\Delta t' = \gamma_v \left[c\Delta t - \frac{v}{c}\Delta x\right] \quad \Rightarrow \quad c\Delta t = \frac{v}{c}\Delta x \tag{2.2.12}$$

Plugging this back into the transformation for the length measured by Bob gives the length contraction:

$$\Delta x' = \gamma_v \left[\Delta x - v\Delta t\right] = \gamma_v \left[\Delta x - v\left(\frac{v}{c}\Delta x\right)\right] = \gamma_v \Delta x \left[1 - \frac{v^2}{c^2}\right] = \frac{\Delta x}{\gamma_v} \tag{2.2.13}$$

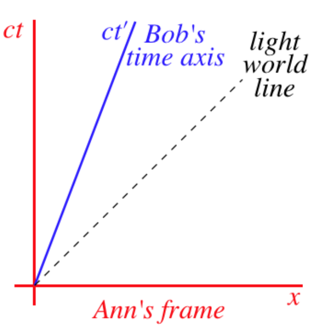
Spacetime Diagrams with Two Observers

Whenever we compare what is seen by two observers in different reference frames, we do so by considering spacetime events, which are common to both. We also looked at how events can be expressed in a spacetime diagram in a specific frame. What we would like to do now is use a *single* spacetime diagram to depict how events are perceived in *both frames at once*.

To do this, we need to first choose one frame (we'll call it Ann's frame)to be the usual spacetime axes, then we need to *plot* the other frame's (Bob's) axes on that same diagram. This is not hard to do now that we have the lorentz transformation. We begin with the same assumption we made for the lorentz transformations – Ann and Bob have a common origin at time $ct = ct' = 0$. This means that their spacetime axes have the same origin.

Next we'll plot Bob's time axis on Ann's spacetime diagram. But wait, what does "plot the time axis" mean? As we know already, even the time axis for Ann is a world line for an object at rest at the origin in Ann's frame. Bob's time axis represents the same thing – the world line of an object at rest at Bob's origin. We are assuming that Bob's frame is moving in the $+x$ direction at a constant speed relative to Ann's frame, so Bob's time axis is a straight line starting at the origin with a slope of $\frac{c}{v}$.

Figure 2.2.1 – Bob's Time Axis



There is another way we can confirm this graph of Bob's time axis. The definition of anyone's time axis is the collection of points where the position on the x-axis remains a constant value of zero. If we plug $x' = 0$ into the Lorentz transformation equations, we get:

$$0=\gamma_v\left[x-\left(\frac{v}{c}\right)ct\right] \quad \Rightarrow \quad ct=\left(\frac{c}{v}\right)x \tag{2.2.14}$$

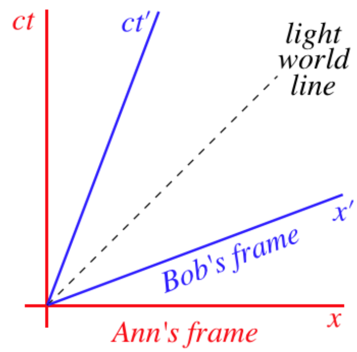
So Bob's ct' axis is a linear plot on Ann's axes (which are ct vs. x) intersecting with the origin with slope $\frac{c}{v}$.

Okay, how do we plot Bob's x' axis on Ann's spacetime diagram? The x'-axis isn't a world line (it has a slope less than that of light), so we can't use that argument again. We can throw together some logic for what the x'-axis should be, but it is simpler to use the lorentz transformation as we did in our confirmation of the time axis above. The x'-axis is defined as the set of points for which the time value ct' remains a constant zero:

$$0=\gamma_v\left[ct-\left(\frac{v}{c}\right)x\right] \quad \Rightarrow \quad ct=\left(\frac{v}{c}\right)x \tag{2.2.15}$$

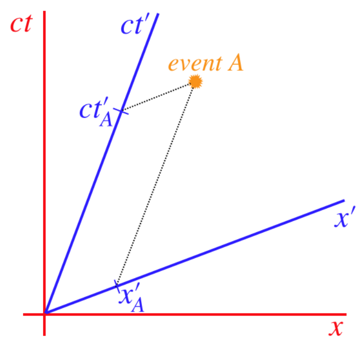
So Bob's x'-axis is a line on Ann's spacetime diagram that passes through the origin with a slope of $\frac{v}{c}$. This is simply the line that is symmetrically-placed across the time world line from the ct'-axis:

Figure 2.2.2 – Bob's Frame in Ann's Spacetime Diagram



We know how to get the time and space coordinates using Ann's axes, but how do we do it with Bob, when the axes are not mutually perpendicular? Well, the points that make up Bob's time axis are the world line of an object at rest at the origin. An object at rest at $x'=1m$ will be moving at the same speed in Ann's frame as the object at the origin, so it's world line will be parallel to the ct' axis. This means that we can find the x' position of an event by tracing a path back to the x'-axis *parallel to the ct' axis*. Similarly, the ct' coordinate can be found by tracing a straight line from the event parallel to the x' axis back to the ct' axis.

Figure 2.2.3 – An Event Recorded in Bob's Frame



It is important to note that the scaling of the unprimed and primed axes in this diagram are not equal. Keeping in mind that the ct'-axis is the world line of the primed frame's origin represented in the unprimed frame, then if we pick a point on the ct'-axis, the value of Δs from the origin to that point is the same for both frames. This means that the tick marks representing fixed units of distances on the primed frame are farther apart than those on the unprimed frame. Let's look at an example to make this clear:

Suppose the primed frame is moving at $0.8c$ relative to the unprimed frame. Then the slope of the ct'-axis is $\frac{5}{4}$. Let's look at a point on the ct'-axis that is 4 units over on the x-axis from the origin, and 5 units up the ct-axis. This point on the ct'-axis needs to be marked $\Delta s=\sqrt{5^2-4^2}=$ 3 units. So the tick-mark for 3 units on the ct'-axis is farther from the origin than the tick-mark for 5 units on the ct-axis is from the origin.

With this graphical tool, we can demonstrate all the previous results we have obtained – some easily, some with difficulty. Perhaps the easiest is simultaneity. Place any two events that are simultaneous in one of the two frames, and it becomes immediately apparent that they cannot be simultaneous in the other. What is more, it also clear which of the two events comes first in the frame in which they are not simultaneous.

The invariance of the speed of light is another result that is easy to see, at least for light traveling in the $+x$ direction. The world line of a light beam moving in the $+x$ direction must have a slope of $+1$, which needs to split both pairs of axes. For light traveling in the $-x$ direction, it is not immediately obvious that the two axes again share the same world line. But extend the $v=-c$ world line until it hits both axes, and it is easy to see that in both cases it intersects the same values on the ct and x-axes in both frames (it is perpendicular to the $v=+c$ world line, so the symmetry is clear).

2.3: Velocity Addition

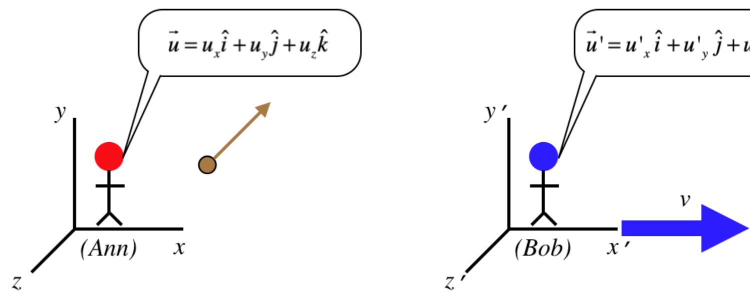
Cosmic Speed Limit

We have already said a few times that nothing we have found makes any sense if ordinary objects can move at or beyond the speed of light. So doesn't the whole theory come crashing down in the following scenario?

Bob is moving past Ann at $0.6c$ in the $+x$-direction, as she fires off her powerful potato gun in the $-x$-direction. Ann measures the speed of the potato in her frame to be $0.8c$. Bob now measures the speed of the potato – doesn't he see it moving at $1.4c$, ruining this whole crazy theory?

This sensible result comes straight from the equally-sensible (and yet wrong) Galilean transformation. Let's consider a case where both Ann and Bob are looking at the same object that is moving relative to both of them (as always, they are moving with a relative speed of v along the x-axis):

Figure 2.3.1 – Ann and Bob Watch the Same Ball



If we use the Galilean transformation to determine how these velocities $\overrightarrow{u}$ and $\overrightarrow{u}'$ relate to each other, we get:

$$\begin{array}{lll} t' = t & \Rightarrow & \dfrac{dt'}{dt} = 1 \\ x' = x - vt & \Rightarrow & u'_x = \dfrac{dx'}{dt'} = \dfrac{dx'}{dt}\dfrac{dt}{dt'} = \left(\dfrac{dx}{dt} - v\right)(1) = u_x - v \\ y' = y & \Rightarrow & u'_y = \dfrac{dy'}{dt'} = \dfrac{dy'}{dt}\dfrac{dt}{dt'} = \left(\dfrac{dy}{dt}\right)(1) = u_y \\ z' = z & \Rightarrow & u'_z = \dfrac{dz'}{dt'} = \dfrac{dz'}{dt}\dfrac{dt}{dt'} = \left(\dfrac{dz}{dt}\right)(1) = u_z \end{array} \tag{2.3.1}$$

So this gives the result we expected before we ever heard of relativity. But now we know that the Galilean transformations are just an approximation of what really describes our universe, and that we actually need to use the lorentz transformation equations instead. So let's follow precisely the method we followed above with the lorentz transformation to find how velocities add in relativity. [*Note that we have to use the γ that goes with the relative velocity of the two frames (since that is what transforms the coordinates), not the gamma related to the relative motion of the ball. This velocity is not changing, so γ_v is treated as a constant in the derivatives to come.*]

$$\begin{array}{llll} ct' = \gamma_v\left[ct - \left(\dfrac{v}{c}\right)x\right] & \Rightarrow & \dfrac{dt'}{dt} = \gamma_v\left[1 - \left(\dfrac{v}{c^2}\right)\dfrac{dx}{dt}\right] & = \gamma_v\left(1 - \dfrac{u_x v}{c^2}\right) \\ x' = \gamma_v(x - vt) & \Rightarrow & u'_x = \dfrac{dx'}{dt'} = \dfrac{dx'}{dt}\dfrac{dt}{dt'} = \gamma_v\left(\dfrac{dx}{dt} - v\right)\left[\gamma_v\left(1 - \dfrac{u_x v}{c^2}\right)\right]^{-1} & = \dfrac{u_x - v}{1 - \frac{u_x v}{c^2}} \\ y' = y & \Rightarrow & u'_y = \dfrac{dy'}{dt'} = \dfrac{dy'}{dt}\dfrac{dt}{dt'} = \left(\dfrac{dy}{dt}\right)\left[\gamma_v\left(1 - \dfrac{u_x v}{c^2}\right)\right]^{-1} & = \dfrac{u_y}{\gamma_v\left(1 - \frac{u_x v}{c^2}\right)} \\ z' = z & \Rightarrow & u'_z = \dfrac{dz'}{dt'} = \dfrac{dz'}{dt}\dfrac{dt}{dt'} = \left(\dfrac{dz}{dt}\right)\left[\gamma_v\left(1 - \dfrac{u_x v}{c^2}\right)\right]^{-1} & = \dfrac{u_z}{\gamma_v\left(1 - \frac{u_x v}{c^2}\right)} \end{array} \tag{2.3.2}$$

Note that the *signs* of the components of $\overrightarrow{u}$ are important here. When $u_x > 0$, the x-component of the ball's velocity measured in Ann's frame is in the same direction as Bob is moving relative to Ann, and Bob sees the velocity of the ball as a little bit faster than the Galilean transformation predicts, because the denominator of the velocity transformation is less than 1. On the other hand, if the ball's velocity has an x-component in Ann's frame that is opposite to Bob's direction of motion ($u_x < 0$), then the velocities add, but the sum is not as great as is predicted by the Galilean transformation, because the denominator is greater than 1.

So let's try the example that started this discussion. Putting in the velocity of the potato in Ann's frame and Bob's relative velcoity with Ann gives the speed measured by Bob:

$$u'_x = \frac{u_x - v}{1 - \frac{u_x v}{c^2}} = \frac{-0.8c - 0.6c}{1 - \frac{(-0.8c)(0.6c)}{c^2}} = \frac{-1.4c}{1 + 0.48} = -0.946c \tag{2.3.3}$$

The cosmic speed limit is obeyed!

Example 2.3.1

*Here's another idea to bring relativity to its knees: Maybe we can't get a speed greater than c for a moving **object** by stacking one velocity on top of another because neither of them can be going a speed c to begin with, but what if Ann shines a light so that the light's speed adds to Bob's speed? Won't Bob measure the light going faster than c?*

Solution

Well, this clearly violates the postulate of relativity, but let's use this idea to check our velocity formula anyway. Putting $u_x = \pm c$ into the velocity transformation indeed gives us the right answer:

$$u'_x = \frac{\pm c - v}{1 - \frac{(\pm c)(v)}{c^2}} = \pm c$$

Velocity Vectors

We have focused on the effects of velocity addition along the direction of motion, but perhaps a result we didn't see coming was that relative motion along the x-direction also affects the relationship between the two observers' measurements of velocities in directions perpendicular to the relative motion. This comes about because of the time dilation effect – if time is passing more slowly in another frame than in your own, then things are moving more slowly in those perpendicular directions. For example, if Bob sees time passing more slowly for Ann as she moves by, then when she drops her pencil, it will take longer to get to the floor from Bob's perspective than from Ann's perspective, so for Bob the pencil is moving more slowly, even though it is moving in a direction perpendicular to the relative motion.

Suppose as Bob zooms by, Ann fires a laser in the $+y$ direction. What does Bob see for the laser's speed and direction? Well, he had better see a speed of c, and thanks to the velocity transformation, he does:

$$\left.\begin{array}{lll} u_x = 0 & \Rightarrow & u'_x = \dfrac{0 - v}{1 - 0} = -v \\ u_y = c & \Rightarrow & u'_y = \dfrac{c}{\gamma_v (1 - 0)} = \sqrt{c^2 - v^2} \end{array}\right\} \left|\overrightarrow{u}'\right| = \sqrt{u'^2_x + u'^2_y} = c \tag{2.3.4}$$

Bob and Ann don't agree on the *direction* however. While Ann sees the laser going in the $+y$ direction, Bob sees an angle with his y'-axis:

$$\sin\theta = \frac{v}{c} \quad \Rightarrow \quad \theta = \sin^{-1}\left(\frac{v}{c}\right) \tag{2.3.5}$$

2.4: Momentum Conservation

Another Thought Experiment

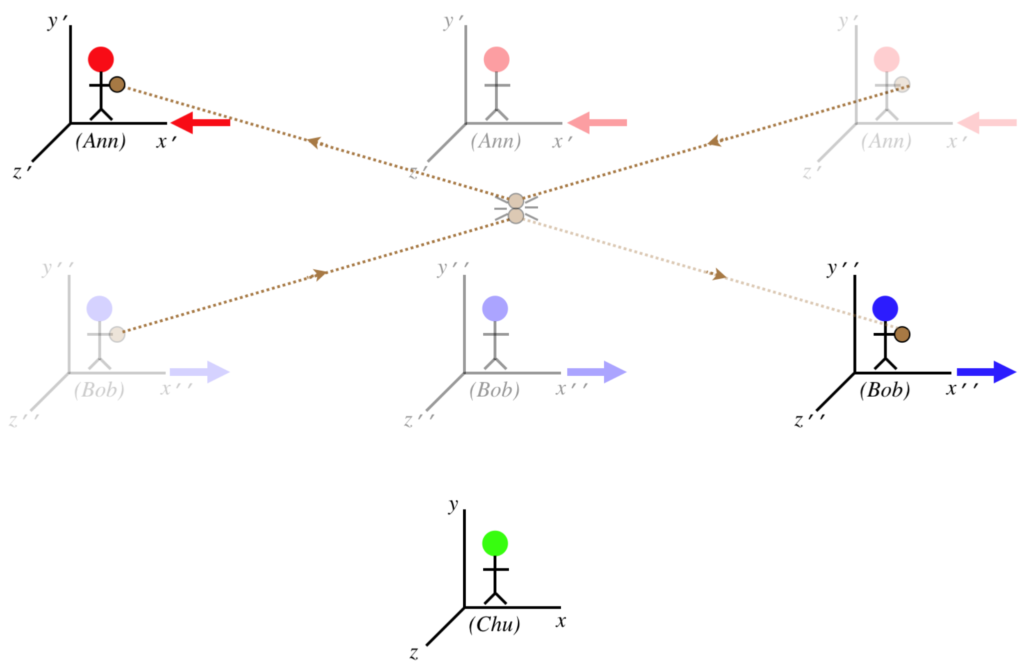
We have completed our exploration of the fundamentals of relativistic kinematics and its consequences. Now it is time to examine the consequences of the relativity principle in the area of dynamics. Our first clue that something needs to be done differently comes from the basis of all dynamics: Newton's laws of motion. In particular, the second law looks like:

$$\overrightarrow{F}_{net}=\frac{d\overrightarrow{p}_{cm}}{dt} \tag{2.4.1}$$

The obvious question is, with respect to what time measurement is this momentum changing? This is actually a very difficult problem to deal with in relativity in that it is difficult to apply the relativity principle, so we will instead look at a consequence of this law – momentum conservation. If someone in one frame observes a collision of two objects and declares that momentum is conserved, then an observer in another frame watching the same collision should conclude the same thing. Momentum conservation is among the most cherished principles of physics, and if an experiment could be performed where two inertial observers do not agree that it is upheld, then that would cause problems for the relativity principle. This calls for a thought experiment!

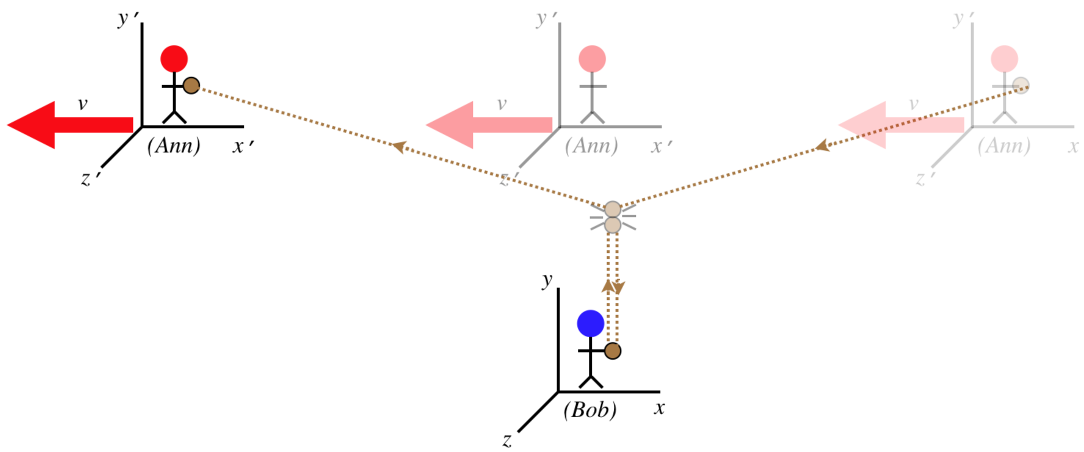
We have three players involved here (Ann, Bob, and Chu), and we are initially viewing the activities from Chu's perspective, who sees Ann moving in the $-x$-direction and Bob moving in the $+x$-direction at equal speeds. Ann and Bob both throw identical balls in a direction that is (from their own perspectives) along their y axes, at speeds they each measure to be u. The balls collide elastically with each other, and return to where they started.

Figure 2.4.1 – Chu Observes Collision



From Chu's perspective, everything is completely symmetric, and so he concludes that the momentum (and most notably, the y-component of the momentum) in this collision is conserved. Okay, let's confirm that all observers agree upon this basic physics principle by looking at the very same collision from Bob's perspective. To Bob, his ball's motion parallels his y-axis, while Ann's has an x-component (equal to her frame's relative motion).

Figure 2.4.2 – Bob Observes Collision



Bob now compares the y-components of velocity of the two balls. We already stated that Bob measures the velocity of his ball to be u, and Ann measures velocity of her ball to be the same, but because of the strangeness of velocity, neither agrees that the *other* ball is moving that fast. Bob says that the y-component of velocity of Ann's ball is (u_x and u_y are the x and y components of Ann's ball, respectively, according to Ann):

$$u'_y \left(\text{Ann's ball}\right) = \frac{u_y}{\gamma_v\left(1-\frac{u_x v}{c^2}\right)} = \frac{u}{\gamma_v\left(1-0\right)} = \frac{u}{\gamma_v} \tag{2.4.2}$$

This is *smaller* than the vertical speed that Bob measures for his own ball, which means that Ann's ball (which has the same mass as his ball) had to have less momentum going into the collision than his ball. But he sees is own ball come back at the same speed that it left, and the collision is elastic (his ball isn't any warmer), and these are properties of a head-on collision with an identical ball moving at the same speed in the opposite direction if momentum is conserved. So something is wrong here.

A New-and-Improved Momentum

There are only three possible conclusions we can draw from the discrepancy shown in this thought experiment:

- Momentum conservation is not a fundamental principal of physics, since one observer measuring no change in momentum (in the example above, Chu) for a closed system does not ensure that every other observer (Ann and Bob) gets that same result.
- Mass is measured differently in different reference frames. If Bob measures the mass of Ann's ball to be greater than his own, then that could compensate for the lower velocity so that their two balls once again have equal y-components of momenta.
- Our definition of momentum, while useful for low velocities, needs a facelift to handle relativistic speeds.

For physicists, stopping at the first of these was never an option – conservation principles and momentum in particular had been revered for far too long. In the early years (and for quite some time afterward), the second option was the accepted explanation, and it works fine. In modern times, however, the physics community has instead embraced the third option – we prefer to characterize mass as an invariant quantity that is inherent to matter, and simply admit that our original definition of momentum was insufficient.

Deriving the correct form of the momentum is challenging until we get to 4-vectors in a future chapter, so right now we will just be given this formula and confirm that it works for the Ann/Bob/Chu thought experiment above. For an object moving at a speed u relative to an observer, the momentum of the object in the frame of that observer is defined to be:

$$\overrightarrow{p} \equiv m\gamma_u \overrightarrow{u} \tag{2.4.3}$$

Alert

Up to now, we have usually dealt with γ_v that relates the frames of two observers, whose relative speed we define as 'v,' but here the γ_u relates the frame of the observer to that of the moving object, not another observer. As we will be relating the momenta in frames of observers, it will be important to keep straight the difference between γ_v and γ_u. As one example, if we have two observers, the γ_v that relates their frames are the same for both of them, but the γ_u that one observer uses for a moving object is not the same as the γ'_u used by the other observer for the same moving object, since $u \neq u'$.

As a first check, it is clear that at slow velocities, momentum reduces to our usual definition of momentum, since for $u << c$, $\gamma_u \rightarrow 1$.

Proving that this form of momentum is conserved in the thought experiment above requires careful accounting of velocities, as there are many involved here. In an effort to not have to carry primes on all our variables through the calculation, we will look at this collision from Ann's perspective – clearly the collision has all the same features for her that it has for Bob. We will use the subscript 'A' whenever referring to a quantity specifically related to Ann's ball, and use a 'B' for Bob's ball. We seek to write all of these quantities in terms of the value u, the speed that each measures for their own ball; and v, the relative speed of their two frames.

We need the γ's for the two balls from Ann's perspective in order to compare momenta in her frame. The γ for her ball is easy, as she sees it moving with a speed of u:

$$\overrightarrow{p}_A = -m\,\gamma_{uA}\,u_A\,\hat{j} = -\frac{m\,u}{\sqrt{1-\frac{u^2}{c^2}}}\,\hat{j} \tag{2.4.4}$$

When Ann looks at the ball thrown by Bob, she sees that it has an x-component equal to $+v$, and its y-component is determined from the velocity addition formula. Bob sees no x' component for his ball's velocity $u'_{Bx}=0$, and he measures the y' component for his ball's velocity to be $u'_{By}=u$, so using the velocity addition formula to determine the two components of velocity of Bob's ball in Ann's frame gives:

$$\begin{aligned} u_{Bx} &= \frac{u'_{Bx}+v}{1+\frac{u'_{Bx}v}{c^2}} = \frac{0+v}{1+0} = v \\ u_{By} &= \frac{u'_{By}}{\gamma_v\left(1+\frac{u'_{Bx}v}{c^2}\right)} = \frac{u}{\gamma_v\,(1+0)} = \frac{u}{\gamma_v} \end{aligned} \tag{2.4.5}$$

We now need to construct the γ_u for Bob's ball as seen by Ann, which means we first need to construct the square of the speed for Bob's ball according to Ann:

$$u_B^2 = u_{Bx}^2 + u_{By}^2 = v^2 + \frac{u^2}{\gamma_v^2} \quad\Rightarrow\quad \gamma_{uB} = \frac{1}{\sqrt{1-\frac{u_B^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}-\frac{u^2}{\gamma_v^2c^2}}} \tag{2.4.6}$$

Plugging this in gives the y-component of momentum of Bob's ball as measured by Ann, which we can then compare to the momentum of Ann's ball (which is entirely in the y-direction):

$$p_{By} = \gamma_{uB}\;m\;u_{By} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}-\frac{u^2}{\gamma_v^2c^2}}}\;m\;\frac{u}{\gamma_v} \tag{2.4.7}$$

Showing that momentum is conserved with its new definition is now a matter of showing that this equals the magnitude of Equation 2.4.4:

$$\frac{1}{\sqrt{1-\frac{v^2}{c^2}-\frac{u^2}{\gamma_v^2c^2}}}\;\cancel{m}\;\frac{\cancel{u}}{\gamma_v} = \frac{\cancel{m}\,\cancel{u}}{\sqrt{1-\frac{u^2}{c^2}}} \tag{2.4.8}$$

Inverting and squaring both sides of the equation and simplifying completes the proof:

$$\left(1-\frac{v^2}{c^2}-\frac{u^2}{\gamma_v^2c^2}\right)\gamma_v^2 = 1-\frac{u^2}{c^2} \quad\Rightarrow\quad \overset{1}{\cancel{\left(1-\frac{v^2}{c^2}\right)\gamma_v^2}} - \frac{u^2}{c^2} = 1-\frac{u^2}{c^2} \tag{2.4.9}$$

Due to the symmetry of the two cases, it should be clear that Bob will get the same result, and from Chu's perspective, the new version of momentum changes the magnitude of the momenta of both balls equally, so he will naturally again witness momentum conservation. Indeed, with this new definition of momentum, every frame will agree that it is conserved before and after the collision.

2.5: Energy Conservation

One Last Thought Experiment

With an understanding of relativistic momentum conservation now firmly in hand, we will have a look at an inelastic collision. When we first encountered these in Physics 9HA, we said that no energy was actually "lost," it was just converted into another form (thermal). Given that thermal energy is at its core "microscopically mechanical," this assessment of it changing form is really just a shortcut, and in fact the only things that change in terms of energy is how it is packaged. For example, we could model a simple (perfectly) inelastic collision between two equal masses this way:

Figure 2.5.1 – A Simple Model of an Inelastic Collision

In this collision, if we can see what is going-on inside the boxed system after the collision, we can account for all of the incoming energy – part of it goes to the kinetic energy of the boxed system, and part of it to the potential and kinetic energy associated with the oscillation of the two particles. If we can't see what's going on, then we can only see the kinetic energy of the boxed system, and we call the leftover energy "internal energy" within the boxed system.

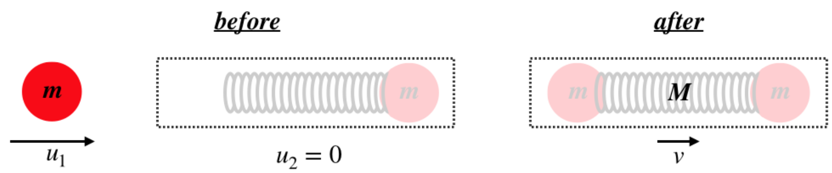
In Physics 9HA, for the non-relativistic case of a collision of this kind, we showed that the fraction of the initial kinetic energy the box system has after the collision is given by:

$$\frac{KE_{after}}{KE_{before}} = \frac{m_1}{m_1 + m_2} = \frac{1}{2} \tag{2.5.1}$$

Put another way, for this case (non-relativistically, when the masses are equal), the energy contained in the oscillation of the two masses equals the kinetic energy of the box system after the collision.

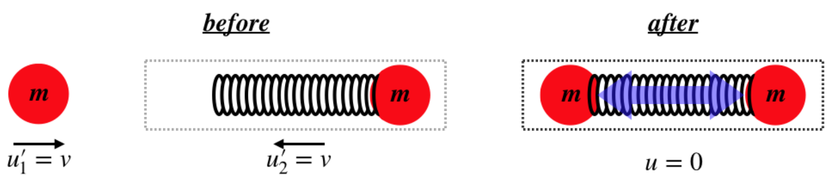
Let's see how all this works out for relativity with our new definition of momentum. We will watch this collision from two different perspectives. The first is Ann, who sees the collision from the perspective shown above, and who cannot see inside the boxed system, so she doesn't even know the system's mass after the collision (she calls it 'M'). Here is the before/after diagram she uses for momentum conservation:

Figure 2.5.2 – Ann's Before/After Diagram for a Perfectly Inelastic Collision



Bob will view the very same collision from another frame that is moving to the right with a speed of v, which is in the rest frame of the system after the collision. Unlike Ann, we'll say that he is able to see what is going on inside the box. With the two parts having equal mass, and coming to rest after the collision, he naturally must see both halves moving at the same speed, so from his perspective, the collision looks like this:

Figure 2.5.3 – Bob's Before/After Diagram for a Perfectly Inelastic Collision



Okay, so let's invoke relativistic momentum conservation for Ann. If she uses this to determine the unknown mass M, she finds:

$$\gamma_{u_1}\, m\, u_1 + 0 = \gamma_v\, M\, v \quad \Rightarrow \quad M = \left(\frac{\gamma_{u_1}}{\gamma_v}\right)\left(\frac{u_1}{v}\right)\, m \tag{2.5.2}$$

Next all we have to do is relate u_1 and v to each other using velocity addition:

$$u_1=\frac{u_1'+v}{1+\frac{u_1'v}{c^2}}=\frac{2v}{1+\frac{v^2}{c^2}}\quad\Rightarrow\quad\frac{u_1}{v}=\frac{2c^2}{c^2+v^2} \tag{2.5.3}$$

Writing γ_{u_1} in terms of v:

$$\gamma_{u_1}=\frac{1}{\sqrt{1-\frac{u_1^2}{c^2}}}=\frac{1}{\sqrt{1-\left(\frac{u_1}{v}\cdot\frac{v}{c}\right)^2}}=\frac{1}{\sqrt{1-\left(\frac{2c^2}{c^2+v^2}\cdot\frac{v}{c}\right)^2}}=\frac{c^2+v^2}{\sqrt{c^4-2v^2c^2+v^4}}=\frac{c^2+v^2}{c^2-v^2} \tag{2.5.4}$$

Plugging everything back into Equation 2.5.2 reveals the mass that Ann measures for the combined system:

$$M=\left(\frac{\frac{c^2+v^2}{c^2-v^2}}{\gamma_v}\right)\left(\frac{2c^2}{c^2+v^2}\right)m=2\gamma_v m \tag{2.5.5}$$

Wait, Ann finds that the mass of the combined system is actually *larger* than $2m$? This doesn't seem to agree with what Bob knows about the combined system. If the two masses were just held together, then Bob would be certain that the combined system would have a mass of $2m$, but perhaps there is something about the fact that the masses are oscillating on the spring that adds to the mass of the system?

Perhaps we can get a clue about what is going on by looking at the mass discrepancy in the case of our familiar slow-moving world. The apparent additional mass is:

$$\Delta m=M-2m=M-\frac{M}{\gamma_v}=M\left(1-\frac{1}{\gamma_v}\right)=M\left(1-\sqrt{1-\frac{v^2}{c^2}}\right) \tag{2.5.6}$$

Now using the usual $u<<c$ approximation:

$$\sqrt{1-\delta}\approx 1-\frac{1}{2}\delta\quad\Rightarrow\quad 1-\sqrt{1-\frac{v^2}{c^2}}\approx 1-\left(1-\frac{1}{2}\frac{v^2}{c^2}\right)=\frac{1}{2}\frac{v^2}{c^2}\quad\Rightarrow\quad\Delta m\approx\frac{1}{2}M\frac{v^2}{c^2} \tag{2.5.7}$$

For this low-velocity case of two equal masses, we said above that the kinetic energy of the box system after the collision equals the internal energy contained in the oscillations, we therefore have:

$$\text{internal energy}=\frac{1}{2}Mv^2=\Delta mc^2 \tag{2.5.8}$$

It appears that the energy that starts as kinetic and becomes internal due to an inelastic collision is manifested – according to relativity – as an increase in the mass of the system where the internal energy is contained, with a conversion factor of c^2.

Total, Kinetic, and Rest Energy

What constitutes "internal energy" is determined by what we define as a system: Just aggregate a group of particles, and that group's collective mass is not simply the sum of the masses of the particles in the group, but must also include the mass that is equivalent to the energy of all their internal motion and interactions according to $E=mc^2$. This famous equation is known as *mass-energy equivalence*, and it has interesting implications. For example, if we make an object hotter, then it contains more internal energy and therefore has more mass than when it is cooler.

In the example above, let's suppose Bob can't see inside the box. As we have said before, observers agree on masses, so he and Ann agree that the box has a mass of M. The box isn't moving in Bob's frame, so internal energy is the *only* energy the box has. The energy of a system measured in its rest frame is called the *rest energy*, and it comes from the system's total mass and mass-energy equivalence:

$$E_{rest}=Mc^2 \tag{2.5.9}$$

If the energy of a system is instead measured in a frame in which it is not at rest, then there is a kinetic energy component that needs to be added to the rest energy to get the total energy. Naturally the total energy of a given system will be greater in frames in which the velocity of the system is greater. We can write the total energy as an unknown function of the velocity of the system in the frame, multiplied by the rest energy:

$$E_{tot}=f(u)\,mc^2 \tag{2.5.10}$$

To obtain this function, let's look at the collision above from Bob's perspective. The energy is conserved, and at the end it is just the rest energy. Before the collision, the two particles have equal total energies whose sum is the final energy:

$$f(v)\,mc^2 + f(v)\,mc^2 = Mc^2 \quad \Rightarrow \quad f(v) = \frac{M}{2m} \quad \Rightarrow \quad f(v) = \gamma_v \tag{2.5.11}$$

The final equality comes courtesy of Equation 2.5.5. So we conclude that the total energy of an object with mass m moving at a speed of u is given by:

$$E_{tot} = \gamma_u mc^2 \tag{2.5.12}$$

Example 2.5.1

Show that energy is conserved for the collision above when measured in Ann's frame.

Solution

The total energy of the system before the collision comes in two pieces – the total energy of the incoming mass, and the rest energy of the target mass. After the collision the system's energy consists of the total energy of the moving total mass. We seek to show that these are equal:

$$\gamma_{u_1} mc^2 + mc^2 \stackrel{?}{=} \gamma_v Mc^2 \quad \Rightarrow \quad \gamma_{u_1} + 1 \stackrel{?}{=} \gamma_v \frac{M}{m}$$

Now substitute for $\frac{M}{m}$ using Equation 2.5.5, giving:

$$\gamma_{u_1} + 1 \stackrel{?}{=} 2\gamma_v^2$$

Now use Equation 2.5.4 to put everything in terms of v and c, and do the algebra:

$$\begin{aligned} \frac{c^2+v^2}{c^2-v^2} + 1 &= \frac{2}{1-\frac{v^2}{c^2}} \\ \frac{c^2+v^2+c^2-v^2}{c^2-v^2} &= \frac{2c^2}{c^2-v^2} \end{aligned}$$

With the total energy and the rest energy of a system now in hand, it is easy to define the kinetic energy as the difference of the two:

$$KE = (\gamma_u - 1)\,mc^2 \tag{2.5.13}$$

Example 2.5.2

Show that the relativistic kinetic energy is consistent with the non-relativistic definition of kinetic energy for speeds much less than c.

Solution

Whenever we see the phrase "speeds much less than c," we immediately think of expanding γ to first order in $\frac{v^2}{c^2}$, as we did in Equation 2.5.7:

$$KE = [\gamma_u - 1]\,mc^2 = \left[\left(1-\frac{u^2}{c^2}\right)^{-\frac{1}{2}} - 1\right] mc^2 \approx \left[\left(1+\frac{1}{2}\frac{u^2}{c^2}\right) - 1\right] mc^2 = \frac{1}{2}mu^2$$

Combining Energy and Momentum

Back in Physics 9HA, we found a very useful formula that relates kinetic energy to momentum. It's clear that the same formula does not work for relativity:

$$\begin{aligned} KE &= (\gamma_u - 1)\,mc^2 \\ \frac{p^2}{2m} &= \tfrac{1}{2m}(\gamma_u mu)^2 = \tfrac{1}{2}\gamma_u^2 mu^2 \end{aligned} \tag{2.5.14}$$

This doesn't mean that there is no formula that relates these two quantities. Indeed:

$$E^2 = \gamma_u^2 m^2c^4 = \frac{c^2}{c^2-u^2}m^2c^4 = \left(\frac{u^2}{c^2-u^2}+1\right)m^2c^4 = \frac{c^2}{c^2-u^2}m^2u^2c^2 + m^2c^4 = \gamma_u^2 m^2u^2c^2 + m^2c^4 = p^2c^2 + m^2c^4 \tag{2.5.15}$$

So the alternative ways of writing the total energy are:

$$E=\gamma_u mc^2=\sqrt{p^2c^2+m^2c^4} \tag{2.5.16}$$

Massless Particles

With the γ_u multiplying mc^2 in the energy equation, we have another reason to insist that the speed of light is unobtainable – for a system to attain the speed of light, it would need to acquire infinite energy. But if this is true, does that mean that light has infinite energy? Of course not – we can measure the energy in light by absorbing it in matter and measuring the temperature change of the matter. So then how does light get away with moving at the cosmic speed limit? The answer is that while γ_u for light goes to infinity, the mass of a light "particle" (called a *photon*) turns out to be zero. The product of these two numbers turns out to result in a finite value.

Using the other energy equation tells us even more. Setting the mass equal to zero gives us a very simple relationship between the energy of a photon and its momentum:

$$E_{photon}=pc \tag{2.5.17}$$

So yes, light has both energy *and* momentum. Again, it might seem strange that something without mass can have momentum, but with γ_u exploding to infinity and the mass vanishing, this is again possible. The difference between light and matter in this regard is that photons don't have any rest energy – all of the energy comes from its momentum.

CHAPTER OVERVIEW

3: Spacetime

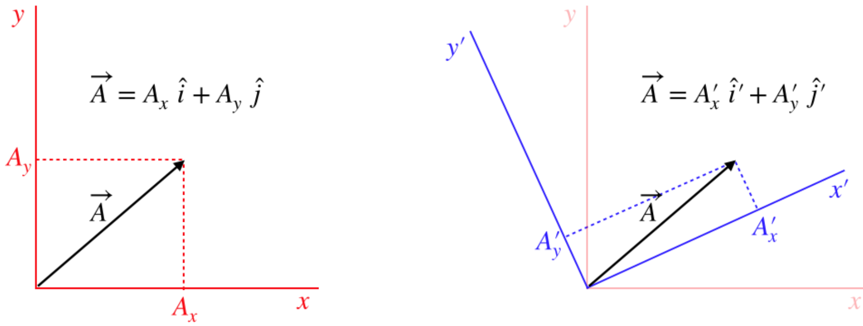
3.1: Vector Rotations

Active vs. Passive Rotations

We take a moment now away from relativity to explore more of the formalism related to vectors. In particular, we will be looking at changes to vectors that result in changing their directions while leaving their magnitudes intact – rotations.

In our dealings with vectors in Physics 9HA, we primarily dealt with vectors in terms of their components in a coordinate system, and we will continue with that practice here. The interesting thing about doing this is that this coordinate representation depends upon the choice of the coordinate axes along which the components are defined. That is, we can look at the same vector in two different coordinate systems, giving us different components for the same vector.

Figure 3.1.1 – Viewing the Same Vector in Two Coordinate Systems



Alert

It is worth emphasizing that we can only compare two vectors using their components if these components are measured on the same axes. Two vectors that are identical (i.e. represent the same physical quantity, giving us all the same experimental results, etc.) can be expressed entirely differently by their coordinates when different coordinate frames are used, so one must take care not to attribute too much "reality" to the components.

Looking at the diagram above, it's clear that the vector $\overrightarrow{A}$ does not change orientation from one frame to the other – it is the same physical quantity whose components are measured with different axes – but if the people using these coordinate systems were to compare the angles they see this vector making with their respective x-axes, they would not agree. This might prompt them to say that the vector is "rotated" when going from one frame to the other. It isn't, of course, but we can express this mathematically as a coordinate transformation that looks like a rotation nonetheless. Such a transformation is called a *passive rotation*, to emphasize that the vector itself is not rotated.

Of course, vectors can also have their actual directions changed as well, such as a rock's velocity vector changing direction as it swings around a circle while tied to a string. These direction changes are due to some physical process, rather than a simple change of measurement perspective. When measured in the same coordinate system, these physical rotations are called *active rotations*.

Because both are typically expressed in terms of components, it is common to confuse active rotations for passive ones, so it is a good idea to keep in mind what is responsible for the rotation to keep them separate. In our study of relativity, our primary focus is on viewing physical quantities from different perspectives, so we will mainly deal with passive transformations in the sections to come. In order to avoid having to append the words "active" or "passive" every time we use the word "rotation," from this point forward we will assume that the rotation is passive unless explicitly stated otherwise.

Invariants

Rotations are defined by the fact that the magnitude of the vector doesn't change. We therefore declare the *vector magnitude to be an invariant with respect to rotations*. There are other invariants as well. The most trivial of these is the independent scalar (like the number 2), which has no connection to the orientation of spatial axes. One might be inspired to say, "The vector magnitude is a scalar, so of course it is an invariant!", but this logic is flawed. The x-component of a vector is also a scalar, but it changes when the coordinate system is rotated.

A vector-related quantity whose invariance may not be readily-apparent is the scalar product of two vectors. As stated above, we can't declare it an invariant simply because it is a scalar, and in fact when expressed in terms of coordinates, it is by no means obvious that the answer comes out the same in both coordinate systems:

$$\overrightarrow{A}\cdot\overrightarrow{B}=A_xB_x+A_yB_y+A_zB_z\overset{?}{=}A'_xB'_x+A'_yB'_y+A'_zB'_z \tag{3.1.1}$$

We will see how to write the primed coordinates in terms of the unprimed coordinates shortly, but we can resolve this issue without doing so by considering the other form of the dot product:

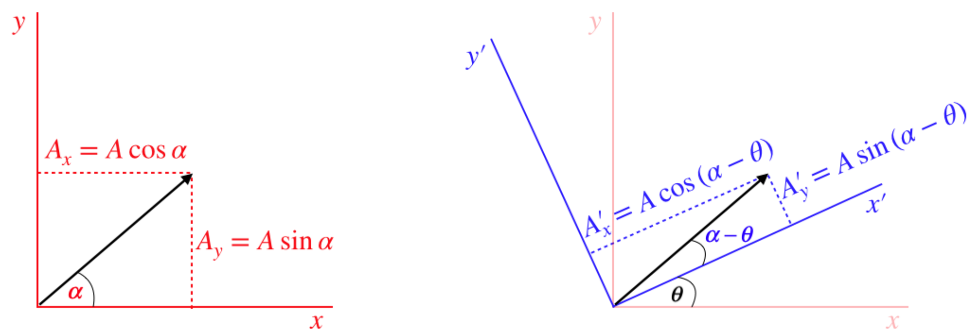
$$\overrightarrow{A}\cdot\overrightarrow{B}=AB\cos\theta \tag{3.1.2}$$

Looking at this representation of the scalar product, we see that it depends upon the magnitudes of the two vectors, which we already know are invariant. Our passive rotation does not rotate the actual directions of either vector, so although the angles these vectors make with the axes change when the axes are rotated, *the angle these two vectors make with each other doesn't change*. With A, B, and θ all left unchanged by the rotation of the coordinate system, the scalar product remains unchanged, and is therefore invariant. Note that the magnitude-squared of a vector is simply a scalar product of the vector with itself ($\overrightarrow{A}\cdot\overrightarrow{A}=A^2$), so these two invariants are consistent with each other.

Transforming Vectors Between Rotated Frames

In keeping with our quest of expressing measurements made in one frame in terms of measurements made in another, we will examine the mathematics associated with translating between components of vectors measured by two coordinate systems rotated with respect to each other (called a *rotational transformation* between coordinate systems). We start by labeling quantities in Figure 3.1.1:

Figure 3.1.2 – Rotational Transformation



The primed coordinate axes have been rotated an amount θ in the positive (counterclockwise) direction relative to the unprimed coordinate axes. We can now write the components in the unprimed frame in terms of the angles α (the angle the vector makes in the unprimed frame) and θ using trigonometric identities for the difference of two angles:

$$\begin{aligned} A'_x &= A\left[\cos(\alpha-\theta)\right]=A\left[\cos\alpha\cos\theta+\sin\alpha\sin\theta\right]=\left[A\cos\alpha\right]\cos\theta+\left[A\sin\alpha\right]\sin\theta \\ &= A_x\cos\theta+A_y\sin\theta \\ A'_y &= A\left[\sin(\alpha-\theta)\right]=A\left[-\cos\alpha\sin\theta+\sin\alpha\cos\theta\right]=-\left[A\cos\alpha\right]\sin\theta+\left[A\sin\alpha\right]\cos\theta \\ &= -A_x\sin\theta+A_y\cos\theta \end{aligned} \tag{3.1.3}$$

To get the inverse transformation, we simply note that the red axes are rotated by the same angle θ relative to the blue axes, but in the opposite direction (clockwise), so simply making the change $\theta\rightarrow-\theta$ does the trick.

Example 3.1.1

Show the rotational transformation maintains the invariance of:

a. *the length of the vector*
b. *the scalar product of two vectors*

Solution

a. Plugging the primed components into the pythagorean theorem and transforming them into unprimed components gives:

$$\begin{aligned} A^2 &= A'^2_x + A'^2_y \\ &= \left(A_x\cos\theta + A_y\sin\theta\right)^2 + \left(-A_x\sin\theta + A_y\cos\theta\right)^2 \\ &= \left(A_x^2\cos^2\theta + \cancel{2A_xA_y\cos\theta\sin\theta} + A_y^2\sin^2\theta\right) + \left(A_x^2\sin^2\theta - \cancel{2A_xA_y\sin\theta\cos\theta} + A_y^2\cos^2\theta\right) \\ &= A_x^2\left(\cos^2\theta + \sin^2\theta\right) + A_y^2\left(\sin^2\theta + \cos^2\theta\right) \\ &= A_x^2 + A_y^2 \end{aligned}$$

b. Plugging the primed components into the scalar product and transforming them into unprimed components gives:

$$\begin{aligned} \overrightarrow{A}\cdot\overrightarrow{B} &= A'_xB'_x + A'_yB'_y \\ &= \left(A_x\cos\theta + A_y\sin\theta\right)\left(B_x\cos\theta + B_y\sin\theta\right) + \left(-A_x\sin\theta + A_y\cos\theta\right)\left(-B_x\sin\theta + B_y\cos\theta\right) \\ &= \left(A_xB_x\cos^2\theta + \cancel{A_xB_y\cos\theta\sin\theta} + \cancel{A_yB_x\sin\theta\cos\theta} + A_yB_y\sin^2\theta\right) \\ &\quad + \left(A_xB_x\sin^2\theta - \cancel{A_xB_y\sin\theta\cos\theta} - \cancel{A_yB_x\cos\theta\sin\theta} + A_yB_y\cos^2\theta\right) \\ &= A_xB_x\left(\cos^2\theta + \sin^2\theta\right) + A_yB_y\left(\sin^2\theta + \cos^2\theta\right) \\ &= A_xB_x + A_yB_y \end{aligned}$$

A Bit About Matrices

Back in 9HA, we had an extremely brief encounter with column matrices as a tool for organizing the components of a vector. We will find these increasingly useful in our study of physics, starting with our current discussion of rotational transformations. By filling the rows of a column matrix with the components of a vector, we take for granted that each row represents one of the three axes. If we change our axes (say by rotating them), then each row means something different. So we need to come up with some mathematical procedure for changing the meaning of the rows. That is, we need to generate this change:

$$\begin{pmatrix} \text{component measured along } x\text{-axis} \\ \text{component measured along } y\text{-axis} \\ \text{component measured along } z\text{-axis} \end{pmatrix} \rightarrow \text{rotate axes} \rightarrow \begin{pmatrix} \text{component measured along } x'\text{-axis} \\ \text{component measured along } y'\text{-axis} \\ \text{component measured along } z'\text{-axis} \end{pmatrix} \tag{3.1.4}$$

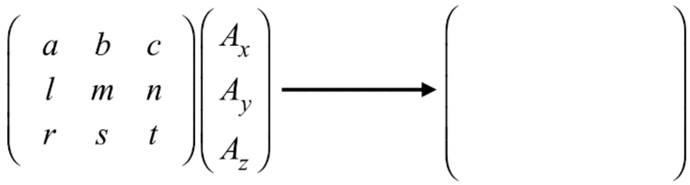
Note that this process represents a *passive* rotation, which means that although the two matrices have different entries, they *represent the same vector*. It's just that the matrix entries are associated with different axes. As we saw above, a component in one set of axes is expressed as a combination of more than one component in the other set of axes. We achieve this mathematically through *matrix multiplication* of a square matrix with the column matrix representing the vector:

$$\begin{pmatrix} A'_x \\ A'_y \\ A'_z \end{pmatrix} = \begin{pmatrix} a & b & c \\ l & m & n \\ r & s & t \end{pmatrix}\begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \tag{3.1.5}$$

The way this multiplication works is this:

- grab the top row of the square matrix and rotate it clockwise by 90 degrees
- align the entries with those of the column matrix, and multiply corresponding entries
- add all of these products together
- place this sum at the top of a new column matrix
- repeat this process by moving to each subsequent row of the matrix, placing the resulting sum in the next row of the new column matrix

Figure 3.1.3 – Matrix Multiplication



Note that this only works when the number of columns of the square matrix matches the number of rows of the column matrix. Indeed more generally an m-by-n matrix (written "$m \times n$", where $m =$ number of rows, $n =$ number of columns) can only multiply (from the left) an $n \times r$ matrix (where r is any positive integer), and the result is an $m \times r$ matrix. This is why transforming a column vector into another column vector requires a square matrix.

Let's put the rotation transformations in Equation 3.1.3 into matrix form. Noting that this rotation is around the z axis, so that the z coordinates don't change, we have:

$$\begin{pmatrix} A'_x \\ A'_y \\ A'_z \end{pmatrix} = \begin{pmatrix} +\cos\theta & +\sin\theta & 0 \\ -\sin\theta & +\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \tag{3.1.6}$$

The matrix that performs this rotation transformation between coordinate axes is called a *rotation matrix*. Keep in mind that this matrix simply helps us express the components of the same vector in a different set of coordinate axes that have (in this case) been rotated counterclockwise around the z-axis by an angle θ.

It's also useful to point out that the scalar product of two vectors can easily be expressed in terms of matrices if the first vector is expressed as a row (1×3) matrix, the second as a column (3×1) matrix, resulting in a scalar (1×1 matrix):

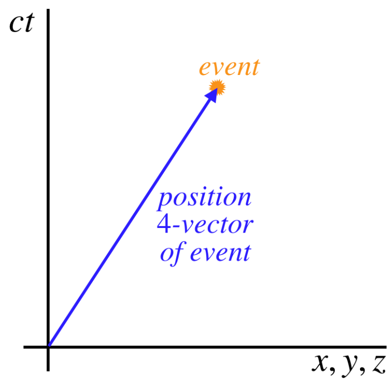
$$\overrightarrow{A} \cdot \overrightarrow{B} = \begin{pmatrix} A_x & A_y & A_z \end{pmatrix} \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} = A_x B_x + A_y B_y + A_z B_z \tag{3.1.7}$$

3.2: Position 4-Vector

World Lines as Vectors

World lines are collections of a continuous set of events in spacetime. Consider a straight world line that stretches from the origin to an event. This line has a magnitude (the spacetime interval between the origin and the event) and a direction (starting at the origin and ending at the event). We can therefore officially declare it to be a vector, and since it points from the origin and has units of length, it is a position vector. Clearly the magnitude of this position vector is quite different from others we have seen, thanks to the Minkowski "adjustment." Because this vector has four components (three for space and one for time), it is called the *position 4-vector*.

Figure 3.2.1 – Position 4-Vector of An Event



Traditionally, the components of this vector are expressed in a column matrix, with the time component being the first entry. [*The entries of ct, x, y and z are typically numbered 0, 1, 2, and 3, respectively.*] Another convention commonly followed is denoting a 4-vector with an upper-case letter (sometimes boldfaced) and no arrow over it, while 3-vectors (those we are accustomed to that exist in space without the time fourth dimension) are lower-case letters with the arrow above them. Here is a summary of conventions we will use here:

$$\text{position 4-vector:} \quad X \leftrightarrow \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \text{ or } \begin{pmatrix} ct \\ \overrightarrow{r} \end{pmatrix} \tag{3.2.1}$$

Alert

A note about the use of the double arrow rather than an equal sign in the expression above... The 4-vector of an event is an object that can be described by any of an infinite number of column matrices, depending upon the frame in which the components are measured. Therefore to call this vector "equal" to one specific set of components is not entirely accurate without somehow acknowledging the frame that that particular matrix represents. The double-arrow is a harmless-but-effective reminder of this important distinction between the 4-vector itself and its representation in a specific frame of reference. It is not "wrong" to use an equal sign (the distinction between the vector and its components is still implied), but we will use the double arrow here for pedagogical purposes.

Magnitude of the Position 4-Vector

Given what we already know about the length of world lines, we can make an important statement about the magnitude of the position 4-vector – it is equal to Δs, which an invariant across inertial frames with coincident origins. This should not surprise us, as the event that it points to is agreed-upon by everyone.

We know that the square of the length of a vector is equal to the scalar product of the vector with itself, and applying that principle here shows us that we need to redefine how we do the scalar product. The usual "multiply like components and add the products" doesn't work, since the signs aren't right. Writing the scalar product in terms of matrices shows this:

$$X \cdot X = \begin{pmatrix} ct & x & y & z \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = c^2t^2 + x^2 + y^2 + z^2 \neq \Delta s^2 \tag{3.2.2}$$

We therefore make an adjustment to the definition of the spacetime scalar product of two vectors in matrix form by placing a square matrix between the vectors that takes care of the *Minkowski metric*:

$$X \cdot X = \begin{pmatrix} ct & x & y & z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = c^2t^2 - x^2 - y^2 - z^2 = \Delta s^2 \tag{3.2.3}$$

[*The 'Δ' in Δs here represents the change from the origin, which is why there are no Δ's on the coordinates.*]

Relating Two Frames

After our previous section on rotations, and showing here how the magnitude of the position 4-vector is invariant between frames of reference, it should be clear where we are going next. We already know that the orientation of a straight world line between two events is different in the spacetime diagrams of two different inertial observers (though it is a straight line in both frames), so we now make the claim that a Lorentz transformation linking two frames results in a sort of 4-dimensional "rotation" of the position vector.

Expressing this in matrix form is a simple matter of repeating the process we followed to get matrix Equation 3.1.6 from Equations 3.1.3, this time starting with Equations 2.2.9. The result is:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma_v & -\frac{v}{c}\gamma_v & 0 & 0 \\ -\frac{v}{c}\gamma_v & \gamma_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{3.2.4}$$

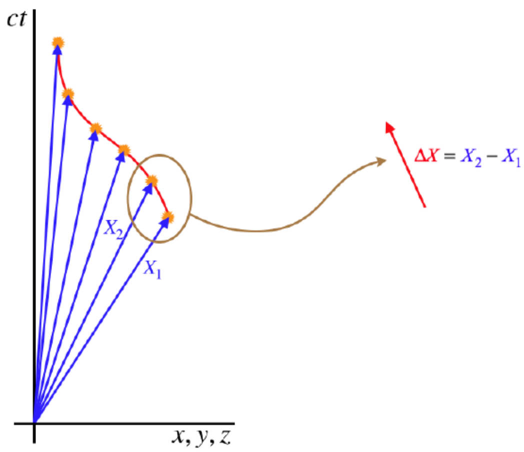
Thanks to the weirdness associated with including the time coordinate, this isn't exactly a rotation, as can be seen by comparing this equation to Equation 3.1.6, which has the two sine functions with different signs, while the corresponding elements in Lorentz transformation matrix have the same sign (not to mention the fact that the sine and cosine can never exceed a value of 1, while the corresponding elements in the Lorentz transformation matrix can and do exceed this value. Apparently "rotating" spatial coordinates into time coordinates (and vice-versa) is analogous to, but not quite the same as, rotating spatial coordinates into other spatial coordinates.

3.3: Velocity and Acceleration 4-Vectors

Calculus of 4-Vectors

There is not a lot we can do with just the position 4-vector. Just as we did in 9HA, we need to use the position vector to construct a velocity vector. We do it pretty much the same way, though as we will see, there is one additional detail that arises in relativity. An object that moves through spacetime is tracked by its position vector at each moment (spacetime event):

Figure 3.3.1 – Position 4-Vector Tracks an Object's World Line



The change in the position 4-vector is defined in the same way as it is for position 3-vectors – using tail-to-head addition. As before, we want the instantaneous rate of change of the position vector, so we choose a change that is infinitesimally-small. We then need to divide by an infinitesimally-small time interval, which is where relativity complicates things – which time do we use?

Different observers will have different spacetime diagrams, and will measure different coordinate time intervals between events. The choice is therefore clear – if we want to construct a velocity 4-vector that is universal (the vector, not its components!), we have to use the proper time between the two events. We don't have to worry about there being several proper time intervals between events, because these events are infinitesimally-close, making the change a straight line, even if the world line path is non-inertial. We therefore define the velocity 4-vector as:

$$V \equiv \frac{dX}{d\tau} \tag{3.3.1}$$

This process of constructing new 4-vectors from others by incorporating invariants is our go-to tactic. We can construct the acceleration 4-vector this way, and we will use this method to construct the momentum 4-vector in the next section.

Properties of Velocity 4-Vectors

Looking at how we constructed the velocity 4-vector, we see that the magnitude of the tiny displacement along the world line also happens to be the spacetime interval between the two nearby events. We therefore find that the magnitude of the velocity 4-vector is:

$$|V| = \frac{ds}{d\tau} = c \tag{3.3.2}$$

Thanks to the invariance of the interval and the proper time, every observer agrees on this magnitude. At first this must seem like a very strange result – every object's velocity 4-vector has the same magnitude, no matter what its world line looks like, and that magnitude is the speed of light?! This is only confusing until one gets accustomed to thinking about velocity 4-vectors differently from velocity 3-vectors. To demonstrate this difference and show this property more clearly, let's express the velocity 4-vector in a specific reference frame:

$$V = \frac{dX}{d\tau} \quad \leftrightarrow \quad \frac{d}{d\tau}\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c\frac{dt}{d\tau} \\ \frac{dx}{d\tau} \\ \frac{dy}{d\tau} \\ \frac{dz}{d\tau} \end{pmatrix} \tag{3.3.3}$$

We want to express this 4-vector in terms of quantities measured in the frame, namely the 3-vector velocity $\overrightarrow{u}$ of the object being observed. The derivatives of the positions with respect to the proper time are not the components of the 3-vector velocity – for that we need derivatives with respect to coordinate time. To this end, we use the chain rule and the relation between proper time and coordinate time (time dilation):

$$\frac{dt}{d\tau} = \gamma_u \quad \Rightarrow \quad \frac{dx}{d\tau} = \frac{dx}{dt} \cdot \frac{dt}{d\tau} = u_x \cdot \gamma_u \tag{3.3.4}$$

Putting this into all the components of the velocity 4-vector matrix gives:

$$V \leftrightarrow \gamma_u \begin{pmatrix} c \\ u_x \\ u_y \\ u_z \end{pmatrix} = \gamma_u \begin{pmatrix} c \\ \overrightarrow{u} \end{pmatrix} \tag{3.3.5}$$

Example 3.3.1

Use the velocity 4-vector matrix to show that its magnitude-squared is c^2 two different ways:

a. *Do it for an arbitrary reference frame.*
b. *Pick a convenient reference frame and argue why you can do this.*

Solution

a. The magnitude-squared of any vector is the dot product of that vector with itself, but for 4-vectors we have to be careful to incorporate the Minkowski metric, so following the matrix multiplication shown in Equation 3.2.3:

$$V \cdot V = \begin{pmatrix} \gamma_u c & \gamma_u u_x & \gamma_u u_y & \gamma_u u_z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \gamma_u c \\ \gamma_u u_x \\ \gamma_u u_y \\ \gamma_u z \end{pmatrix} = \gamma_u^2 \left(c^2 - u_x^2 - u_y^2 - u_z^2\right) = \gamma_u^2 \left(1 - \frac{u^2}{c^2}\right) c^2 = c^2 \tag{3.3.6}$$

b. If we choose the rest frame of the object in question, then its 3-velocity is zero, making $\gamma_u = 1$, and leaving the velocity 4-vector with only one component – the time component, equal to c, giving a magnitude squared of c^2. The magnitude of this 4-vector is an invariant, which means that all reference frames will get this same result.

Acceleration 4-Vectors

Knowing that every object's velocity 4-vector has the same magnitude, and that this magnitude remains the same for all time, may inspire us to ask about acceleration. On this count, there are two important considerations: First, objects can clearly change the magnitudes of their velocity 3-vectors (i.e. they can accelerate in the sense that we are used to) – it's just that the time component of their velocity 4-vectors will compensate for these changes such that the 4-vector magnitude remains fixed. Second, just because the magnitude of a velocity 4-vector doesn't change, it doesn't mean that its *direction* (in spacetime) doesn't. This can happen two distinct ways – the velocity 3-vector can change direction, or it can speed-up/slow down (or both, obviously). The first is an rotation in space, and the second is a "rotation" between the space and time components.

We can construct the acceleration 4-vector in the same manner that we constructed the velocity 4-vector – by taking a derivative with respect to proper time. To write this vector in terms of a column matrix gets significantly messier than it was for the velocity, because the derivative will now act on factors of γ_u present in the components of the velocity 4-vector that were not present in the position 4-vector. If we consider acceleration that is only along the direction of motion x (so the object is only speeding-up or slowing-down), then the magnitude of the (3-vector) acceleration is just the derivative of the magnitude of the (3-vector) velocity:

$$\overrightarrow{a} \,||\, \overrightarrow{u} \quad \Rightarrow \quad a = \frac{du}{dt} \tag{3.3.7}$$

This makes the derivative of γ_u equal to:

$$\frac{d\gamma_u}{dt} = \frac{d}{dt}\left(1 - \frac{u^2}{c^2}\right)^{-\frac{1}{2}} = \left(1 - \frac{u^2}{c^2}\right)^{-\frac{3}{2}} \frac{u}{c^2}\frac{du}{dt} = \frac{u}{c^2}\,\gamma_u^3\, a \tag{3.3.8}$$

Before proceeding, let's note the following useful identity:

$$1 + \frac{u^2}{c^2}\,\gamma_u^2 = \gamma_u^2 \tag{3.3.9}$$

Now we apply all of this to the construction of the acceleration 4-vector matrix:

$$A = \frac{dV}{d\tau} = \frac{dt}{d\tau} \cdot \frac{dV}{dt} \quad \leftrightarrow \quad \frac{dt}{d\tau} \cdot \frac{d}{dt} \begin{pmatrix} \gamma_u c \\ \gamma_u u \\ 0 \\ 0 \end{pmatrix} = \gamma_u \begin{pmatrix} \frac{u}{c}\,\gamma_u^3\, a \\ \frac{u^2}{c^2}\,\gamma_u^3\, a + \gamma_u a \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{u}{c}\,\gamma_u^4\, a \\ \gamma_u^2 \left(\frac{u^2}{c^2}\,\gamma_u^2 + 1\right) a \\ 0 \\ 0 \end{pmatrix} = \gamma_u^4\, a \begin{pmatrix} \frac{u}{c} \\ 1 \\ 0 \\ 0 \end{pmatrix} \tag{3.3.10}$$

Example 3.3.2

Derive the 4-vector acceleration components in terms of the 3-vector velocity and 3-vector acceleration for the more general case when these two 3-vectors are not parallel. [Note: You will need to write the u^2 that appears in γ_u as a dot product of the 3-vector velocity with itself, and then make use of the product rule on the dot product.]

Solution

Start with the definition of the acceleration 4-vector as the derivative of the 4-vector velocity with respect to proper time:

$$A=\frac{dV}{d\tau}=\frac{dt}{d\tau}\frac{dV}{dt}=\gamma_u\frac{dV}{dt}$$

Now we need to perform the derivative. The 4-vector velocity includes a factor of γ_u, which includes the quantity $u^2=\vec{u}\cdot\vec{u}$, so let's do this little pieces at a time. The derivative of u^2 gives us the derivative of γ_u:

$$\frac{d}{dt}u^2=\frac{d}{dt}\vec{u}\cdot\vec{u}=2\vec{u}\cdot\frac{d\vec{u}}{dt}=2\vec{u}\cdot\vec{a}\quad\Rightarrow\quad\frac{d}{dt}\gamma_u=\frac{d}{dt}\left(1-\frac{u^2}{c^2}\right)^{-\frac{1}{2}}=\frac{1}{2c^2}\left(1-\frac{u^2}{c^2}\right)^{-\frac{3}{2}}[2\vec{u}\cdot\vec{a}]=\gamma_u^3\frac{\vec{u}\cdot\vec{a}}{c^2}$$

Now we put this result into the full derivative of V:

$$A=\gamma_u\frac{d}{dt}\left[\gamma_u\left(\begin{array}{c}c\\\vec{u}\end{array}\right)\right]=\gamma_u\left[\frac{d\gamma_u}{dt}\left(\begin{array}{c}c\\\vec{u}\end{array}\right)+\gamma_u\frac{d}{dt}\left(\begin{array}{c}c\\\vec{u}\end{array}\right)\right]=\gamma_u^4\frac{\vec{u}\cdot\vec{a}}{c^2}\left(\begin{array}{c}c\\\vec{u}\end{array}\right)+\gamma_u^2\left(\begin{array}{c}0\\\vec{a}\end{array}\right)=\left(\begin{array}{c}\gamma_u^4\frac{(\vec{u}\cdot\vec{a})}{c}\\\frac{1}{c^2}\gamma_u^4\left(\vec{u}\cdot\vec{a}\right)\vec{u}+\gamma_u^2\vec{a}\end{array}\right)$$

It is left as an exercise for the reader to show that this general result reduces to the result above when $\vec{u}$ is parallel to $\vec{a}$. [Note that this means $\vec{u}\cdot\vec{a}=ua$.]

Example 3.3.3

Show that the 4-vector acceleration is always perpendicular to the 4-vector velocity.

Solution

*There are no fewer than **three** good ways to solve this. Two of them require clever and powerful relativistic arguments, while the third method consists of brute force algebra. All three of these are satisfying in their own way...*

Method 1: using the invariant magnitude of V

The magnitude-squared of the 4-vector velocity is c^2, so:

$$0=\frac{d}{d\tau}\left(c^2\right)=\frac{d}{d\tau}(V\cdot V)=2\frac{dV}{d\tau}\cdot V=A\cdot V$$

Naturally, two vectors with a zero dot product are orthogonal.

Method 2: using the invariance of the dot product and a convenient frame

The dot product is invariant with respect to choice of frame, so if it is zero in one frame, it vanishes in all frames. Choose the rest frame, $\vec{u}=0$, $\gamma_u=1$:

$$V\cdot A=\left(\begin{array}{c}c\\\vec{0}\end{array}\right)\cdot\left(\begin{array}{c}0\\\vec{a}\end{array}\right)=0$$

Method 3: brute force

It is left as an exercise for the reader to perform the dot product between the 4-vector velocity given in Equation 3.3.5 and the 4-vector acceleration given in the result of Example 3.3.2. It is important to remember, however, that the Minkowski metric must be used in this dot product. That is, the product of the time components are multiplied by +1, and the products of the spatial components are multiplied by -1 before these are added together in the dot product.

Let's step back for a moment and consider the case of two observers: One is accelerating at a constant rate along the x-axis, while the other remains in an inertial frame. The person who is accelerating *knows* that they are doing so (they can do a test to see that they are not in an inertial frame), so how does the measurement of acceleration made by one observer relate to the measurement made by the other?

At first we might expect both observers to measure the same (3-vector) acceleration, but there is a major problem with this. There is no physical law that says that the observer in the accelerated frame can't keep accelerating indefinitely at the same rate – they just need to keep the rocket thrusters set at the same level for as long as they like. But the other observer cannot see this occur, or after a long enough period of time the speed of the other frame relative to theirs will exceed c, since $u(t)=at+u_o$. So the observer in the inertial frame must witness gradually *decreasing* acceleration while the accelerated observer observes constant acceleration. We can show this in a couple of ways.

First, we're not really equipped to talk about non-inertial frames, so let's change the situation to two inertial observers, each looking at the same object that is accelerating through space. One of these observers (we'll call it the primed frame), at the moment of observation, happens to be (momentarily) at rest relative to the object, while the other observer (the unprimed frame) sees the primed frame (and therefore momentarily the object) moving in the $+x$-direction at a speed u. The claim above is that these two observers cannot agree upon the 3-vector acceleration. In the primed frame, we have $u=0$ and $\gamma_u=1$, which gives the following components of 4-vector acceleration:

$$A \overset{u=0}{\leftrightarrow} \begin{pmatrix} 0 \\ a' \\ 0 \\ 0 \end{pmatrix} \tag{3.3.11}$$

Now use the fact that the magnitude-squared of this 4-vector is an invariant to compare the (3-vector) accelerations measured in the two frames. In the rest frame this is easy to compute:

$$A \cdot A = -a'^2 \tag{3.3.12}$$

In the unprimed frame we have:

$$A \cdot A = \gamma_u^8 \; a^2 \left(\frac{u^2}{c^2} - 1\right) = -\gamma_u^6 \; a^2 \tag{3.3.13}$$

Applying the invariance of the magnitude of 4-vectors means we can set these equal from the two frames, giving the simple result:

$$a' = \gamma_u^3 \; a \tag{3.3.14}$$

As expected, the magnitude of the acceleration measured in the inertial frame (a) is less than what is measured in the rest frame (a').

A second way to check this is to perform the Lorentz transformation on the 4-vector acceleration in one frame to get the 4-vector acceleration in the other, and then check the 3-vector transformation:

$$\gamma_u^4 \; a \begin{pmatrix} \frac{u}{c} \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma_u & \frac{u}{c}\gamma_u & 0 & 0 \\ \frac{u}{c}\gamma_u & \gamma_u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ a' \\ 0 \\ 0 \end{pmatrix} \;\;\Rightarrow\;\; \gamma_u^3 \; a = a' \tag{3.3.15}$$

Example 3.3.4

Bob moves at a constant speed in a circle around Ann, who is in an inertial frame. Use the result of Example 3.3.2 to derive the relationship between the magnitude of the acceleration Bob that measures in his frame to the magnitude measured by Ann.

Solution

The answer comes out immediately by setting $\overrightarrow{u} \cdot \overrightarrow{a} = 0$ *, and setting the two acceleration 4-vector magnitudes equal. With Bob being the primed frame and Ann the unprimed frame, we get:*

$$a' = \gamma_u^2 \; a$$

A Classic Example

With what we understand about acceleration 4-vectors, we can now work out a classic problem solved in most classes in special relativity. It involves the twin paradox, where one twin gets into a spaceship that is constructed to simulate close to the Earth's gravity by accelerating at a constant rate only slightly less than that of earth's gravity: $a = \frac{1.0 \; light-year}{year^2} = 9.5\frac{m}{s^2} \approx g$. This twin takes a round-trip to the nearest star (approximately 4 light years distant) in this ship, increasing its speed for the first half of the trip to the star, decreasing it for the second half so that the ship stops at the star. Then the process is repeated for the return trip, speeding up then slowing down. Our goal is to determine how long this trip takes for the twin in the ship, and how long it takes for the twin on the Earth. [*Notice that this is a much more reasonable set of circumstances than what we have used for the twin paradox before now, where the acceleration was instantaneous.*]

We start by noting that all 4 legs of the trip (speeding up toward the star, slowing down toward the star, speeding up toward the Earth, and slowing down toward the Earth) are all going to give identical results for the times measured in each frame, as they are completely symmetric and the relative directions of the acceleration and velocity 3-vectors are irrelevant as long as they remain parallel. So we only need to do the calculation for the first leg and multiply the result by 4.

The acceleration in the frame of the spaceship is simply g, which means that at the moment during the trip when the spaceship is moving at speed u relative to the Earth, the twin on Earth measures an acceleration of:

$$a = \frac{g}{\gamma_u^3} = g\left(1 - \frac{u^2}{c^2}\right)^{\frac{3}{2}} \tag{3.3.16}$$

This acceleration is the rate at which the speed of the ship is changing, as measured by the earth:

$$\frac{du}{dt} = g\left(1 - \frac{u^2}{c^2}\right)^{\frac{3}{2}} \;\;\Rightarrow\;\; \int\left(1 - \frac{u^2}{c^2}\right)^{-\frac{3}{2}} du = \int g \; dt \tag{3.3.17}$$

These are not difficult integrals to perform. Noting that $u = 0$ at $t = 0$, we have:

$$\frac{u}{\sqrt{1-\frac{u^2}{c^2}}}=gt \tag{3.3.18}$$

Solving for u gives:

$$u(t)=\frac{gt}{\sqrt{1+\frac{g^2t^2}{c^2}}} \tag{3.3.19}$$

Integrating the velocity over the time of the trip gives the distance (halfway to the star):

$$\Delta x=\int_0^T \frac{gt\;dt}{\sqrt{1+\frac{g^2t^2}{c^2}}}=\frac{c^2}{g}\left(\sqrt{1+\frac{g^2T^2}{c^2}}-1\right) \quad\Rightarrow\quad T=\sqrt{\frac{2\Delta x}{g}+\frac{\Delta x^2}{c^2}} \tag{3.3.20}$$

Notice that if not for the effect of "slowing acceleration" seen from the earth frame, the time this leg of the trip takes would be found to be simply:

$$\Delta x=\frac{1}{2}gt^2 \quad\Rightarrow\quad t=\frac{2\Delta x}{g} \tag{3.3.21}$$

This only includes the first term in the square root given above. Plugging-in our values of $\Delta x=2\;ly$, $g\approx 1\frac{ly}{year^2}$, and of course $c=1\frac{ly}{year}$, and multiplying by 4 to get the full time, we get the time of the trip measured on the Earth:

$$T=16\sqrt{2}\;years\approx 23\;years \tag{3.3.22}$$

Now we need to calculate the time measured aboard the spaceship. The spaceship measures the proper time, which we can obtain from Equation 1.2.9, now that we know the speed of the ship as a function of time. Again, we compute one of the 4 intervals, and multiply by 4 to get the total time:

$$\Delta\tau=\frac{\Delta s}{c}=4\int_0^{\frac{T}{4}}\sqrt{1-\frac{u(t)^2}{c^2}}\;dt=4\int_0^{\frac{T}{4}}\sqrt{1-\frac{g^2t^2}{c^2+g^2t^2}}\;dt=4\int_0^{\frac{T}{4}}\frac{dt}{\sqrt{1+\frac{g^2}{c^2}\;t^2}} \tag{3.3.23}$$

At last we come upon an integral that is not a single simple substitution away from being solved, so we take the coward's way out and look it up in an integral table:

$$\int\frac{dx}{\sqrt{1+x^2}}=\ln\left[x+\sqrt{1+x^2}\right] \quad\Rightarrow\quad \Delta\tau=4\frac{c}{g}\,\ln\left[\frac{gT}{4c}+\sqrt{1+\frac{g^2T^2}{16c^2}}\right] \tag{3.3.24}$$

[*Note: Most folks use the identity:* $\sinh^{-1}x=\ln\left[x+\sqrt{1+x^2}\right]$ *to reduce the space required to write this solution.*]

Plugging in for T, g and c and noting that $\frac{c}{g}\approx 1\;year$, gives:

$$\Delta\tau=(4\;years)\ln\left[4\sqrt{2}+\sqrt{1+32}\right]=9.7\;years \tag{3.3.25}$$

It takes less than half as much time to make the round trip for the twin aboard the spacetime as elapses on Earth.

3.4: Momentum 4-Vector

Construction of the Momentum 4-Vector

Back when we first discussed momentum, the form of the relativistic momentum was just given and shown to work. Here we will demonstrate the power of 4-vector thinking by deriving the momentum 4-vector, of which the momentum 3-vector is a part. The whole "derivation" consists of one trivial step. We get the momentum 4-vector from the velocity 4-vector by multiplying it by the mass. The mass is an invariant, so this will ensure that the 4-vector we construct has the necessary property – it has a magnitude that is measured to be the same by all observers. We see that making this definition results in momentum components in the spatial dimensions are exactly the "new" momentum we defined previously:

$$P \equiv mV \quad \leftrightarrow \quad \begin{pmatrix} \gamma_u mc \\ \gamma_u mu_x \\ \gamma_u mu_y \\ \gamma_u mu_z \end{pmatrix} = \begin{pmatrix} \gamma_u mc \\ \gamma_u m \overrightarrow{u} \end{pmatrix} \tag{3.4.1}$$

The ct component of this 4-vector looks familiar. It is in fact common to write the components of the momentum 4-vector as:

$$P \quad \leftrightarrow \quad \begin{pmatrix} \frac{E}{c} \\ \overrightarrow{p} \end{pmatrix} \tag{3.4.2}$$

One compelling reason to write the matrix this way is that it works equally-well for photons as massive objects. But it also assists us in seeing the physical interpretation of momentum and energy according to relativity. Just as space and time need to be treated as different parts of the same whole we now call spacetime, so too are energy and momentum linked. We no longer see conservation of momentum and conservation of total energy as separate physical principals – they are just two parts of the single principle of conservation of 4-vector momentum.

Let's compute the magnitude-squared of this 4-vector, and let's do it the "easy way" – by doing it in the rest frame. After all, the magnitude computed in that frame will be the same in every other frame. Setting $\rightarrow u = 0$ and $\gamma_u = 1$ in the components of the momentum 4-vector above leaves only the ct component, making its square the magnitude-squared:

$$P \cdot P = m^2 c^2 \tag{3.4.3}$$

If we now also compute the magnitude-squared the "hard way" by doing it in an arbitrary frame, we get:

$$P \cdot P = \left(\frac{E}{c}\right)^2 - \overrightarrow{p} \cdot \overrightarrow{p} = \frac{E^2}{c^2} - p^2 \tag{3.4.4}$$

Setting the "easy" and "hard" calculations equal gives us a familiar relation:

$$E^2 = p^2 c^2 + m^2 c^4 \tag{3.4.5}$$

Using the Momentum 4-Vector

We use the momentum 4-vector in precisely the same way that we use the momentum 3-vector (and in fact three quarters of this use is exactly this). That is, we add up all of the momenta in a system before a collision (which can include photons as well as massive particles), add it all up afterward, and set them equal to invoke the conservation principle. Of course, the presence of the γ_u's complicates things somewhat, as particles that collide will naturally change frames going from "before" to "after." On the other hand, the method of choosing a simpler frame to work in such as the system's overall rest frame (just as we occasionally employed the center-of-mass frame in the case of non-relativistic collisions) can be quite powerful in such problems. As with the position and velocity 4-vectors, the Lorentz transformation correctly translates the components of the momentum 4-vector between frames.

3.5: General Relativity

The Equivalence Principle

One thing that comes up over and over in special relativity is the role of inertial frames. We talked about accelerated frames as well, but those followed different rules. The most notable of these is the effect that the acceleration of a frame has on the measurement of the proper time. If an observer carries a clock in their non-inertial frame and measures the time between two events that occur at the same place in that frame with it, the time measured is less than when an observer in an inertial frame carries a clock and does the same with those two events.

This phenomenon is reflected in the shape of the world line of the origin of the non-inertial frame in a spacetime diagram in the rest frame of an inertial observer. By contrast, the shape of the world line of the origin of an inertial frame is always straight when drawn in an inertial observer's space time diagram. The intervals defined by all the straight world lines come out the same.

Einstein must have worried if perhaps it is possible for the world line of an inertial frame to be curved in the spacetime diagram of another inertial observer (it is unlikely he worried about it in exactly these terms, but we'll run with this). This would cause problems, because the Lorentz transformation that links inertial frames only rotates world lines, it doesn't cause them to curve. It turns out that Einstein came up with an example that causes this exact problem!

To see how this problem can arise, let's first return to something we studied in Physics 9HA. When we talked about free-fall, we noted that because every object in an enclosed space experiences the same acceleration when that space is in free-fall, an observer in this frame will assume that nothing in the space is experiencing a force. Einstein elevated this to a fundamental principle that sounds very much like the relativity principle, and called this one the *equivalence principle*:

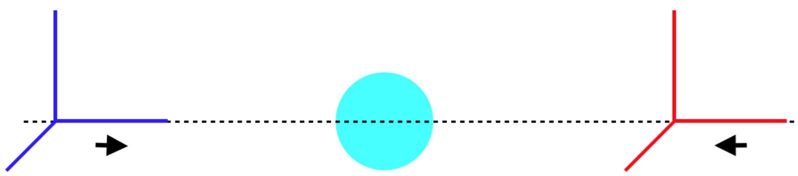
No experiment can be performed that will distinguish a frame in gravitational free-fall from an inertial frame.

The Problem

If this principle is correct, then the following thought experiment raises a problem...

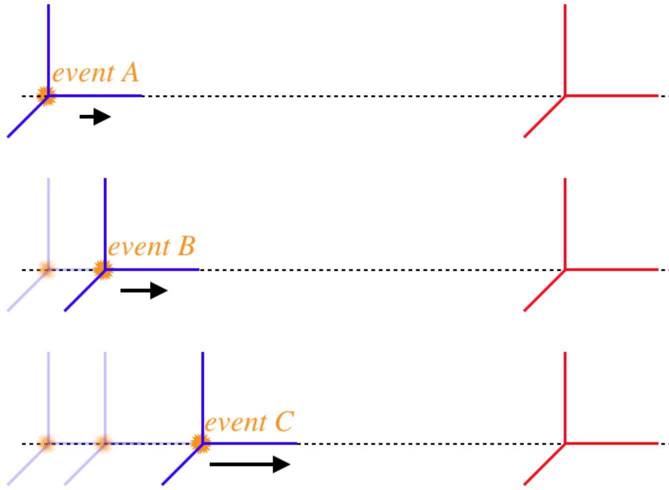
Suppose we have two free-falling frames on opposite sides of a planet. They are moving toward each other in a symmetric fashion, with their common x-axes passing through the center of the planet:

Figure 3.5.1 – Two Free-Falling Frames



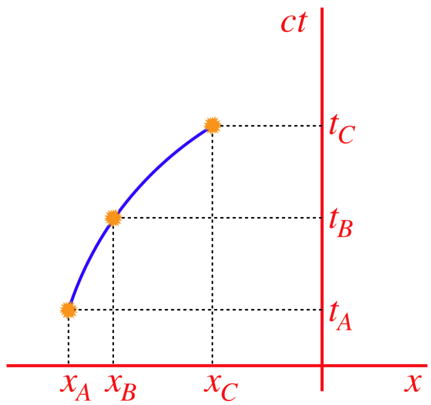
According to the equivalence principle, *both* of these frames are inertial. Now suppose the observer in one of these frames witnesses three events that occur at the origin of the other frame separated by equal coordinate time intervals. The other frame is speeding up relative to the observer's frame, which means that the spatial separation of the first and second events is smaller than that of the second and third events. We'll choose the red frame as the observer frame, and obscure the planet for clarity:

Figure 3.5.2 – Three Events in Free-Falling Frames



To see why this causes us a problem, we now have a look at the spacetime diagram of these events in the observer frame:

Figure 3.5.3 – Spacetime Diagram of Free-Falling Frame



The first thing we notice is that the slope of the world line of the origin of the other frame must be changing, because the three spacetime events are not aligned in the diagram. If we fill-in lots of other events as well, we see that the observed frame has a world line that is a continuous curve that is *not straight*. We have seen this before: Whenever the world line of an object not in an inertial frame is drawn in the spacetime diagram of an inertial frame, we get a curved path through spacetime.

The problem here is that the observed frame *is inertial*, according to the equivalence principle! In fact, we can reverse the roles of these frames – the blue frame can claim to be stationary, and then the spacetime diagram of the world line of the red frame's origin will be curved. Essentially what we have here is two frames that can both claim to be inertial, but when one measures the motion of the other, the second derivative of position with respect to time does not vanish. By all we have studied so far, this constitutes acceleration, which is synonymous with non-inertial.

Einstein's Solution

Clearly gravity is the culprit here – somehow it is redefining how inertial frames relate to one another when they are separated in a region where gravity exists. Our use of Lorentz transformations to related frames needs to be generalized to include cases that involve gravity. This is why what we have studied so far is called *special* relativity. Incorporating gravitation into the same theoretical relativity framework is called *general relativity*.

In the thought experiment above, the blue frame is inertial, and the events all occur at the same position, which means it measures the spacetime interval (the longest proper time interval) between events. For this to be an invariant, it means that the interval measured in the red frame following the path shown in the diagram above must be the same value. But this means that we cannot equate the intervals in the way we have done previously:

$$\Delta s^2 = c^2 \Delta \tau^2 \neq c^2 \Delta t^2 - \Delta x^2 \tag{3.5.1}$$

The reason this is no longer equal is that this relation only holds if the world line is straight (see the discussion of Equation 2.1.7 for a refresher on this). Recall that this same relation applied to the invariance of the lengths of position 4-vectors, and by extension all other 4-vectors as well. We certainly don't want the whole theory to come crashing down, so what is the solution?

Taken in small enough steps, the world line does "look straight," but the machinery we have created around the need for straight world lines requires that it be straight everywhere. So let's instead make the Minkowski metric (which we use in scalar products, including those that define magnitudes of 4-vectors) able to change from place-to-place. That is, instead of the dot product incorporating the simple Minkowski matrix shown in Equation 3.2.3, let's introduce one that can change from one point to the next in spacetime:

$$X \cdot X = \begin{pmatrix} ct & x & y & z \end{pmatrix} \begin{pmatrix} g_t(ct,x) & 0 & 0 & 0 \\ 0 & -g_x(ct,x) & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = g_t c^2 t^2 - g_x x^2 - y^2 - z^2 = \Delta s^2 \tag{3.5.2}$$

This new matrix that defines dot products represents what is called the *metric tensor*. What we have done here is a greatly over-simplified version (so much that we won't be doing any specific mathematics in this area), but the basic idea is correct. In essence, the presence of gravity changes the way intervals are measured, making it different from the usual Minkowski case. The simplest way to describe this change is to say that the *spacetime itself is curved* by the presence of gravity. When there is no gravity present, then the metric coefficients g_t and g_x converge back to 1, and we say that the space is "flat," giving us once again special relativity.

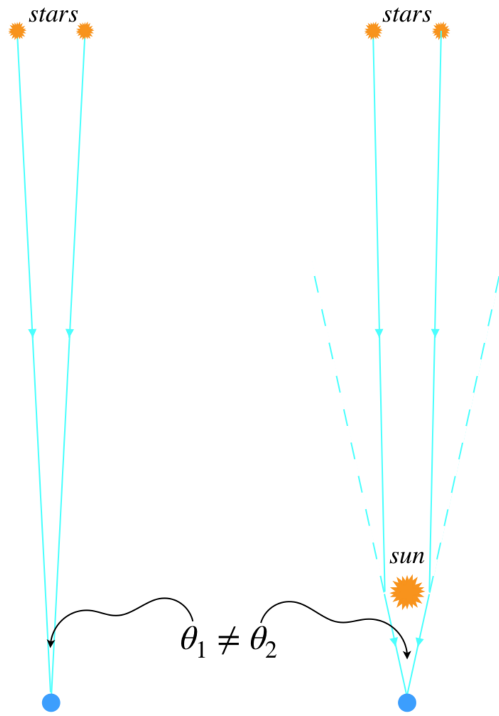
Of course the "hard part" of general relativity is determining how the metric coefficients arise from what we already know is the source of gravity – mass-energy. This is something Einstein spent so many years on (it was 10 years between his publications of special and general relativity), largely because he had to learn a field of mathematics called differential geometry, which essentially deals with curved surfaces. Unfortunately, this lies well beyond the scope of this course.

Effects of Gravity

Now that we know gravity curves spacetime itself, we can discuss a few of its measurable effects. Perhaps the most direct evidence came from the result of the experiment that is generally considered to be the premier confirmation of the theory, and it came more than three years after Einstein's publication of the theory. Light is massless, so from Newton's theory, it should not be effected by gravity. According to general relativity, light should follow a specific world line through spacetime, and gravity bends world lines, so we should be able to measure this bend.

The experiment that made this measurement looked at two stars in the night sky in two ways: The first way was directly – in the night sky. The angular separation of the positions of the stars was noted. The second viewing of these stars was during a full solar eclipse, when they could be seen on opposite sides of the sun. The effect of the sun's gravity on the light coming from these stars would be to bend the spacetime such that the light rays deflect slightly compared to when the sun is not present. This would be manifested as an apparent shift in the angular separation of the stars. Obviously the figure below is not drawn to scale – the dimensions are exaggerated to clearly show the effect.

Figure 3.5.4 – Bending of Starlight



One might try to argue that we can give light an "effective mass" using mass-energy equivalence, then use Newton's theory to predict the effect that gravity has on it, but in fact this approach makes the wrong prediction for the deflection of the light!

Of course gravity also has effects on measured spatial lengths and intervals of time, the latter of which has been measured by atomic clocks placed at different elevations on the Earth (thereby experiencing different gravitational fields), it creates a doppler-like effect on light escaping a gravitating body (called gravitational red-shift), and it makes new predictions for planetary orbits that are different from Newton's (which have also been confirmed), but the relation of these phenomena to the curving of spacetime is not as dramatic as the case of the deflection of starlight.

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: UCD: Physics 9HB – Special Relativity and Thermal/Statistical Physics

Webpages: 27

All licenses found:

- CC BY-SA 4.0: 66.7% (18 pages)
- Undeclared: 33.3% (9 pages)

By Page

- UCD: Physics 9HB – Special Relativity and Thermal/Statistical Physics - *CC BY-SA 4.0*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Foundations of Relativity - *CC BY-SA 4.0*
    - 1.1: The Relativity Principle - *CC BY-SA 4.0*
    - 1.2: The Nature of Time - *CC BY-SA 4.0*
    - 1.3: More Thought Experiments - *CC BY-SA 4.0*
    - 1.4: Paradoxes - *CC BY-SA 4.0*
  - 2: Kinematics and Dynamics - *CC BY-SA 4.0*
    - 2.1: Spacetime Diagrams - *CC BY-SA 4.0*
    - 2.2: Lorentz Transformation - *CC BY-SA 4.0*
    - 2.3: Velocity Addition - *CC BY-SA 4.0*
    - 2.4: Momentum Conservation - *CC BY-SA 4.0*
    - 2.5: Energy Conservation - *CC BY-SA 4.0*
  - 3: Spacetime - *CC BY-SA 4.0*
    - 3.1: Vector Rotations - *CC BY-SA 4.0*
    - 3.2: Position 4-Vector - *CC BY-SA 4.0*
    - 3.3: Velocity and Acceleration 4-Vectors - *CC BY-SA 4.0*
    - 3.4: Momentum 4-Vector - *CC BY-SA 4.0*
    - 3.5: General Relativity - *CC BY-SA 4.0*
  - Back Matter - *Undeclared*
    - Index - *Undeclared*
    - Glossary - *Undeclared*
    - Detailed Licensing - *Undeclared*