Physics 209: Introductory Notes on Relativity

January, 2001

(to be continued)

Last revised 1/14/01

1. The Principle of Relativity

The principle of relativity is an example of an *invariance principle*. There are several such principles. All other things being the same:

1. *Principle of translational invariance in space:* It doesn't matter where you are.

2. *Principle of translational invariance in time:* It doesn't matter when you are.

3. *Principle of rotational invariance:* It doesn't matter how you are oriented.

The principle of relativity fits into the same pattern:

4. *Principle of Relativity:* If you're moving with fixed speed along a straight line, it doesn't matter how fast you're going.

"It doesn't matter" means "the rules for the description of natural phenomena are the same". For example the rule describing the force of gravity between two chunks of matter is the same whether they are in this galaxy or another (*translational invariance in space*). It is also the same today as it was a million years ago (*translational invariance in time*). The law does not work differently depending on whether one chunk is east or north of the other one (*rotational invariance*). Nor does the law have to be changed depending on whether you measure the force between the two chunks in a railroad station, or do the same experiment with the two chunks on a uniformly moving train (*principle of relativity*).

"All other things being the same" raises deep questions. In the case of translational invariance it means that when you move the experiment to a new place or time you have to move everything relevant; in the case of rotational invariance you have to turn everything relevant. In the case of the principle of relativity you have to set everything relevant into motion. If everything relevant turned out to be the entire universe you might wonder whether there was any content to the principle.

One can thus descend immediately into a deep philosophical morass from which some never emerge. We shall not do this. We are interested in how such principles work on the practical level, where they are usually unproblematic. You easily can state a small number of relevant things that have to be the same and that is quite enough. When the principle doesn't work, invariably you discover that you have overlooked something else simple that is also relevant. Not only does that fix things up, but often you learn something new about nature that proves useful in many entirely different contexts. If, for example the stillness of the air was important for the experiment you did in the railroad station, then you had better be sure that when you do the experiment on a uniformly moving train that you do not do it on an open flat car, where there is a wind, so all other relevant things are not the same. You must do it in an enclosed car with the windows shut. If you hadn't realized that the stillness of the air was important in the station, then the apparent failure of the experiment to work the same way on the open flatcar would teach you that it was.

Invariance principles are useful because they permit us to extend our knowledge to new situations. It is on that quite practical level that we shall be interested in the principle of relativity. It tells us that we can't distinguish between a state of rest and a state of uniform motion. Therefore an experiment must give the same result, whether it is performed in a laboratory at rest or in a uniformly moving laboratory. The new situation — doing the experiment in a uniformly moving laboratory — has a result that we can infer from the result we got in the old situation — the laboratory at rest.

It is important both to understand clearly what the principle asserts and to acquire some skill in using it to extend your knowledge from one situation to another. On a deep level one can again get bogged down in subtle questions. What do we mean by rest or by uniform motion? We will again take a practical view. Uniform motion means moving with a fixed speed in a fixed direction.[1] Note that fixed direction is as important as fixed speed: a particle moving with fixed speed on a circle is not moving uniformly. You can clearly tell the difference between being in a plane moving at uniform velocity and being in a plane moving in turbulent air; between being in a car moving at uniform velocity and one that is accelerating or cutting a sharp curve or on a bumpy road or screeching to a halt. But you cannot tell the difference (without looking out the window) between being on a plane flying smoothly through the air at 400 miles per hour and being on a plane that is stationary on the ground.

In working with the principle of relativity the term *frame of reference* is extremely useful. A frame of reference (often simply called a "frame") is the (uniformly moving) system in terms of which you have chosen to describe things. For example a flight attendant walks toward the front of the airplane at 2 miles per hour (mph) in the frame of reference of the airplane. You start at the rear of the plane and want to catch up with him so you walk at 4 mph in the frame of the plane. If the plane is going at 500 mph then in the frame of reference of the ground this would be described by saying that the cabin attendant was

[1] More compactly, moving with a fixed *velocity*. Note that the term "velocity" embraces both speed and direction of motion. Two boats moving 10 miles per hour, one going north and the other east, have the same speed but different velocities. Note also the extremely useful (as we shall see) convention that a *negative* velocity in a given direction means exactly the same thing as the corresponding positive velocity in the opposite direction: -5 mph east is exactly the same as 5 mph west.

moving forward at 502 mph, and you caught up by increasing your speed from 500 mph to 504 mph. One of the many remarkable things about relativity is how much one can learn from considerations of this apparently banal variety.

Another important term is *inertial frame of reference.* "Inertial" means stationary or uniformly moving. A rotating frame of reference is not inertial. Nor is one that oscillates back and forth. We will almost always be interested only in inertial frames of reference and will omit the term "inertial" except when we wish to contrast uniformly moving frames of reference to frames that move nonuniformly.

How do you know that a frame of reference is inertial? This is just another way of posing the deep question of how you know motion is uniform. It would appear that you have to be given at least one inertial frame of reference to begin with, since otherwise you can ask "Moving uniformly with respect to what?" Thus if we know that the frame in which a railroad station stands still is an inertial frame, then the frame of any train moving uniformly through the station is also an inertial frame. But how do we know that the frame of reference of the station is inertial?

Fortunately there is a simple physical test for whether a frame is inertial. In an inertial frame stationary objects on which no forces act remain stationary. It is this failure of a stationary object (you) to remain stationary (you are thrown about in your seat) that lets you know when the plane or car you are riding in (and the frame of reference it defines) is moving uniformly and when it is not. In our cheerfully pragmatic spirit, we will set aside the deep question of how you can know that no forces act. We will be content to stick with our intuitive sense of when the motion of an airplane (train, car) is or is not capable of making us seasick.

When specifying a frame of reference you can sometimes fall into the following trap: suppose you have a ball that (in the frame of reference you are using) is stationary before 12 noon, moves to the right at 3 feet per second (fps) between 12 pm and 1 pm, and to the left at 4 fps after 1 pm. By "the frame of reference of X" (also called the *proper frame* of X) one means the frame in which X is stationary. Now there is no *inertial frame of reference* in which the ball is stationary throughout its history. If you want to identify an *inertial* frame of reference as "the frame of reference of the ball" you must be sure to specify whether you mean the inertial frame in which the ball was stationary before 12, or between 12 and 1, or after 1. There are three different inertial frames that (depending on the time) serve as the frame of reference of the ball. Similarly for the Cannonball Express, which constitutes one inertial frame of reference as it zooms along a straight track at 120 mph from Syracuse to Chicago, and quite another as it zooms along the same track at the same speed on the way back.[2] The frame of reference of an airplane buffeted by high winds may never be inertial. Nor is the frame of reference of the Cannonball Express as it moves smoothly along a curved stretch of track.

Here is another, more subtle trap, that many people (including, I suspect, some physicists) fall into:[3] people sometimes take the principle of relativity to mean, loosely speaking, that the behavior of a uniformly moving object should not depend on how fast it is moving, or, to put it slightly differently, that motion with uniform velocity cannot affect any properties of an object.

This is wrong. The principle of relativity only requires that if an object has certain properties in a frame of reference in which the object is stationary, then if the same object moves uniformly, it will have the same properties *in a frame of reference that moves uniformly with it.* On the other hand the properties of an object moving uniformly past you can certainly differ from the properties the same object has when it is standing still in front of you. To take a trivial example, when the object moves past you it has a non-zero speed; when it is stationary with you its speed is zero.[4]

A more striking example is provided by the so-called Doppler effect: If a yellow light moves away from you at an enormous speed the color you see changes from yellow to red; if it moves toward you at an enormous speed the color changes from yellow to blue. So the color of an object can depend on whether it is moving or at rest (and in what direction it is moving). All the principle of relativity guarantees is that if a light is seen to be yellow when it is stationary, then when it moves with uniform velocity it will still be seen as yellow *by somebody who moves with that same velocity.*

What we shall be almost exclusively interested in are some simple practical applications of the principle of relativity. To apply the principle of relativity it is essential to acquire the ability to visualize how something looks when viewed from different inertial frames of reference. A useful mental device for doing this is to examine how a single set of events would be described by various people moving past them in trains moving uniformly with different speeds.

---

[2] This is, alas, a fictitious train.

[3] I only became fully aware of this trap a few years ago, when reading about some celebrated (but fallacious) objections to relativity by a physicist named Dingle.

[4] You could, of course, object that speed is not a property inherent in an object, but specifies a relation between the object and the frame of reference in which it has that speed. This is fine. The nature of the trap is then that many properties that appear to be inherent turn out, on closer examination, to be relational. We shall see many examples of this.

We will be applying the principle of relativity to learn some quite extraordinary things by examining the same sets of events in different frames of reference. Some of the things we shall learn in this way are so surprising that they are hard to believe at first. You are more likely to conclude that you must have made a mistake in applying the principle of relativity. So it is quite essential to begin by acquiring some skill in using the principle of relativity to learn some things that you might not have known before, which, though not obvious, are also not astonishing. Here is the general trick for doing this:

Take a situation which you don't fully understand. Find a new frame of reference in which you do understand it. Examine it in that new frame of reference. Then translate your understanding in the new frame back into the language of the old one.

Here is a simple example. Newton's first law of motion states that in the absence of an external force a uniformly moving body continues to move uniformly. This law follows from the principle of relativity and a very much simpler law. The simpler law merely states that in the absence of an external force a stationary body continues to remain stationary.

To see how the more general law is a consequence of the simpler one, suppose we only know the simpler law. The principle of relativity tells us that it must be true in all inertial frames of reference. If we want to learn about the subsequent behavior of a ball initially moving at 50 fps in the absence of an external force, all we have to do is find an inertial frame of reference in which we can apply the simpler law. The frame we need is clearly the one that moves at 50 fps in the same direction as the ball, since in that frame of reference the ball is stationary. Putting it more concretely, think of how the ball looks from a train moving at 50 fps alongside it. In the frame of reference of the train the ball is stationary and we can apply the law that in the absence of an external force a stationary body remains stationary. But anything that is stationary in the train frame moves at 50 fps in the frame of reference in which we originally posed the problem. We conclude that since the ball remains stationary in the train frame in the absence of an external force, in the original frame it must continue to move at 50 fps in the absence of an external force.

So starting with the fact that undisturbed stationary objects remain stationary, we have used the principle of relativity to establish the much more general fact that undisturbed uniformly moving objects continue to move with their original velocity.[5] If you

[5] At the risk of complicating something simple, I feel obliged to remark that in reaching this conclusion we have implicitly assumed that if an object is undisturbed in one inertial frame of reference then it is undisturbed in any other inertial frame of reference — i.e. that the condition of no force acting on an object is an *invariant* condition independent of the frame of reference in which the object is described. Since such forces can be associated with jet engines being on or off, springs being compressed or slack, etc., this is a reasonable

already knew Newton's first law you might not be impressed at this line of thought, so let's examine a case where what we learn might not be quite so familiar.

Suppose we have two identical perfectly elastic balls. Identical elastic balls have the property that if you shoot them directly at each other with the same speed, then after they collide each bounces back in the direction it came from with the same speed it had before the collision. Question: What happens if one of the balls is at rest and you shoot the other one directly at it?

There is a long tradition of answering such questions by invoking the conservation of energy and momentum. If you know how to use such conservation laws, you should forget this for now.[6] It is entertaining and instructive that the question can be answered using nothing but the principle of relativity. In learning how to use the principle in this way you will acquire a conceptual skill that will be essential in understanding everything that is to follow. My own feeling is that answering such questions using the principle of relativity provides a deeper insight than answering them by applying conservation laws. Here's how to figure out what happens, using only the principle of relativity:

First draw a picture illustrating the rule you know: when the balls move at each other with equal speeds, they simply rebound with the same speeds. Then draw a picture of the new situation. For concreteness let's take the original speed of the moving ball to be 10 fps. (Once you get good at this business you can simply take it to be a general speed u.) We want to know what goes in the box with the question mark in it.

	Before Collision	**After Collision**
Case 1 (known):	$(X) \rightarrow \ \leftarrow (Y)$	$\leftarrow (X) \qquad (Y) \rightarrow$
Case 2 (unknown):	$(X) \rightarrow \ (Y)$ $10\ fps$	?

To understand what happens in Case 2 hop onto a train moving to the right at 5 fps.[7] Because we are now on a train moving to the right at 5 fps, ball X is moving to the right

assumption.

[6] We shall return to the use of conservation laws in understanding such collisions at the very end of our discussion of the theory of relativity.

[7] Figuring out which train to take is crucial. In the present case we have picked this particular train because it is the one in whose frame the balls are moving with equal and opposite velocities, as we now confirm. Often it is obvious in what frame the unknown situation becomes the known one. Sometimes you have to think about it. At such times

at an additional 5 fps. Since ball Y was stationary before we boarded the train, in the train frame it is moving to the *left* at 5 fps. *Therefore in the frame of this particular train Case 2 (before) is an instance of Case 1 (before).* But the principle of relativity assures us that any experiment we do with the two elastic balls must have the same outcome in any inertial frame of reference. Since the two balls are moving at each other with the same speed in the train frame, after the collision they must bounce away from each other, each still moving at 5 fps in the train frame.

Now all that remains is to translate that answer back to the original frame of reference (which it is convenient to call the track frame). After the collision ball X moves to the left at 5 fps in the train frame, so it must be stationary in the track frame. After the collision ball Y moves to the right at 5 fps in the train frame so it must be moving to the right at 10 fps in the track frame. Therefore the complete picture is this:

	Before Collision	**After Collision**
Case 1 (known):	$(X) \rightarrow \ \leftarrow (Y)$	$\leftarrow (X) \qquad (Y) \rightarrow$
Case 2 (unknown):	$(X) \underset{10\ fps}{\rightarrow} \ (Y)$	$(X) \ \ (Y) \underset{10\ fps}{\rightarrow}$

We have used the principle of relativity to learn something new about identical elastic balls: if one is at rest and the other bumps it head-on, then the moving one comes to a complete stop and the stationary one moves off with the velocity the formerly moving one originally had. This is a fact familiar to all players of billiards, but not many of them realize that it is simply a consequence of the much more obvious fact (less frequently encountered in billiards) that when two balls collide head–on with equal and opposite speeds each bounces back the way it came with its original speed.

As a test to make sure you really understood the above argument, here are two similar questions. They can be answered by a similar application of the principle of relativity. If you understood the argument about the elastic balls, then with a little thought you should be able to answer both of the questions that follow:[8]

trial and error is a useful method. Ask yourself how the balls are described in a frame moving to the right at 1 fps, 2 fps, etc. Frequently the velocity you need then becomes evident.

[8] Conversely, if you don't see how to answer these questions after some thought, then you probably didn't really understand what I was saying about the elastic balls, and should

(1) Two identical sticky balls, depicted in the figure that follows as (X) and (Y), have the property that if they are fired directly at one another with equal speeds, then they stick together upon collision and the resulting compound ball (XY) is stationary. If a sticky ball is fired at 10 fps directly at another identical sticky ball that is stationary and the two stick together, with what speed and in what direction will the compound ball move after the collision?[9]

	Before Collision	**After Collision**
Case 1 (known):	$(X) \rightarrow \; \leftarrow (Y)$	(XY)
Case 2 (unknown):	$(X) \rightarrow \quad (Y)$	?

(2). Suppose we have two elastic balls, but one of them (B) is very big and the other (s) is very small. If the big ball is stationary and the small ball is fired directly at it, the small ball simply bounces back in the direction it came from with the same speed, and the big ball stays at rest. With what speed will each ball move after the collision, if the small ball is stationary and the big ball is fired directly at it with a speed of 15 fps?[10]

	Before Collision	**After Collision**
Case 1 (known):	$(s) \rightarrow \quad (B)$	$\leftarrow (s) \quad (B)$
Case 2 (unknown):	$(s) \quad \leftarrow (B)$	?

In all of these cases you are told how two balls behave under certain conditions and are asked what will happen under a set of conditions that does not fit into the scheme you've been told about. You do this by first finding a frame of reference in which the

think your way through that again.

[9] This can be answered using only the principle of relativity. If you remember how to use conservation of energy and momentum, please forget it.

[10] This one is cunningly devised so that if you try, in spite of my instructions, to solve it using conservation laws instead of the principle of relativity, you will run into difficulties.

new conditions do reduce to the ones you've been told about, then applying the rule you know in that frame, and finally translating the result back into the language of the original frame.

Last revised 1/25/01

2. Nonrelativistic Addition of Velocities

Let us look a little more carefully at the reasoning we used to solve the bouncing ball problem. In addition to using the principle of relativity we also made use of all of the following facts:

1a. If a ball moves down the track at 10 fps in the track frame and a train moves down the track at 5 fps in the track frame then the ball moves down the train[1] at 5 fps in the train frame.

1b. If a ball is stationary in the track frame and the train moves down the track at 5 fps in the track frame then the ball moves up the train at 5 fps in the train frame (or, equivalently, down the train at -5 fps in the train frame).

2a. If a ball moves down the train at 5 fps in the train frame and the train moves down the track at 5 fps in the track frame then the ball moves down the track at 10 fps in the track frame.

2b. If a ball moves up the train at 5 fps in the train frame and the train moves down the track at 5 fps in the track frame then the ball is stationary in the track frame.

We used 1a and 1b to translate the unknown asymmetric collision in the track frame into a known symmetric colision in the train frame. Then we appealed to the principle of relativity, which assures us that the rule[2] about symmetric collisions is valid in any inertial frame of reference. Using this rule we were able to say what happened after the collision in the train frame. Finally we used 2a and 2b to translate the situation after the collision in the train frame back into track-frame language.

As far as we know today, the principle of relativity is indeed valid. But what about assumptions 1a, 1b, 2a, and 2b? How can they be justified? They are all applications of the following rule, which is called the Nonrelativistic Velocity Addition Law:[3] 0 If A, B,

---

[1] Let us agree that "down the train" means towards the front of the train, and "up the train" means towards the rear.

[2] "If two identical elastic balls collide with equal and opposite velocities then after the collision each bounces back in the direction it came from with its original speed."

[3] "Nonrelativistic" is an unfortunate term, but everybody uses it and so shall we. It does *not* mean, as you might think, "in contradiction to the principle of relativity". Unfortunately the body of lore constructed by applying the principle of relativity to certain facts about the speed of light has come to be known as the "Theory of Relativity". The term "nonrelativistic" is invariably used to mean "the way we thought the world behaved before we learned about the theory of relativity". Since (as we shall see) things actually do behave pretty much the way we used to think they did before we learned about the

and C all move with uniform velocity along the same straight line[4] then

$$v_{AC} = v_{AB} + v_{BC}, \tag{2.1}$$

where v_{XY} means "the velocity of X with repect to Y" or, more awkwardly, but more precisely, "the velocity of X in the frame of reference in which Y is stationary."

Before we can check this out, we need to adopt a convention on the direction of motion along which velocities are taken to be positive. We use the same convention in all frames of reference, whether they be the track frame, the frame of train #1, the frame of train #2, etc. Everybody, regardless of what frame of reference he or she uses, agrees on which direction to describe as "down the tracks" and which to describe as "up the tracks". If, for example, the tracks run east-west, everybody can agree that east is down the tracks and west is up the tracks. They also agree to take velocities *in their frame* to be positive or negative, depending on whether the direction of motion *in their frame* is down the tracks or up the tracks. I stress "in their frame". If the train goes down the tracks at 10 fps and the ball goes up the train at 3 fps then in the train frame the ball is going toward the rear of the train and therefore in the direction up the tracks, even though in the track frame the ball is going down the tracks (at 7 fps). So in the train frame the velocity of the ball down the train is −3 fps, while in the track frame it is +7 fps down the tracks.

When we draw pictures of events along the track we shall take the tracks be more or less horizontal and shall take down the tracks to be to the right, and up the tracks to be to the left. The velocities appearing the the velocity addition law (2.1) are velocities in different frames but they must all use the same convention on which direction counts as positive and which as negative. That direction can either be down the tracks or up the tracks, as long as people in all frames use the same convention. Generally we shall take it to be down the tracks — i.e. to the right in the figures we draw.

The four facts mentioned above are all instances of (2.1) with A being the ball, B being the train, and C being the track. Both 1a and 2a assert that the velocity of the ball in the track frame (10 fps) is the velocity of the ball in the train frame (5 fps) plus the velocity of the train in the track frame (5 fps). Both 1b and 2b assert that the the velocity of the ball in the track frame (0 fps) is the velocity of the ball in the train frame (−5 fps) plus the velocity of the train in the track frame (5 fps).

---

theory of relativity, provided all speeds of interest are much less than the speed of light, *nonrelativistic* as used today means precisely *valid to a high degree of accuracy when all speeds are small compared with the speed of light.*

[4] For almost all of the points we shall be making it is enough to consider objects confined to move along a single direction — which we shall often taken to be the direction of a long straight railroad track. There are only two possible directions of motion along such a track.

Usually it is easier just to reason one's way to the answer without having to invoke the abstract form (2.1) of the addition law. Later, however, we will find that the addition law (2.1) is not exactly correct, but only valid to an extremely high degree of accuracy when all speeds are small compared with the speed of light. When velocities comparable to the speed of light enter the picture we shall have to use a modified form of (2.1) and it is then necessary to use the modified addition law because common-sense reasoning no longer gives the right answer.

One important fact, which remains valid even when speeds are comparable to that of light, is that

$$v_{XY} = -v_{YX}. \tag{2.2}$$

If X moves with a certain speed with respect to Y, then Y moves with that same speed with respect to X, but in the opposite direction. This (fairly obvious) relation can be deduced from the general rule (2.1), if you note that because (2.1) holds for arbitrary choices of A, B, and C, it must continue to hold if we interchange B and C:

$$v_{AB} = v_{AC} + v_{CB}. \tag{2.3}$$

Now if we subtract v_{CB} from both sides of (2.3) then we have

$$v_{AB} - v_{CB} = v_{AC}. \tag{2.4}$$

But the forms for v_{AB} given in (2.4) and (2.1) are consistent only if

$$v_{BC} = -v_{CB}. \tag{2.5}$$

There is an even simpler way to see that (2.2) is a consequence of (2.1), that also introduces some terminology we shall subsequently find useful. If an object moves uniformly then there is an inertial frame of reference — the one that has the same velocity as the object — in which the object is stationary. That frame is called the *rest frame* or *proper frame* of the object.[5] Consider the nonrelativistic velocity addition law (2.1) in the special case in which C and A are identical, so that v_{AC} becomes v_{AA}, the velocity of A in the frame in which A is staionary, i.e. in its rest frame. The velocity of A in its rest frame is 0, so we have

$$0 = v_{AA} = v_{AB} + v_{BA}, \tag{2.6}$$

[5] Warning: If an object is not moving uniformly then there is no inertial frame in which it remains stationary at all times. A frame of reference in which a non-uniformly moving object remains stationary must be a non-inertial frame.

and therefore

$$v_{AB} = -v_{BA}. \tag{2.7}$$

How would one go about justifying rule (2.1) to a stubborn person who did not find it obvious? Consider the instance of it provided by 2a. There we can argue as follows: If the ball moves down the train[6] at 5 fps then in one second the ball gets 5 feet further down the train. And if the train moves down the track[7] at 5 fps then in one second the train gets 5 feet further down the track. So in one second the ball gets 10 feet further down the track — the 5 it gains on the train and the additional 5 the train gains on the track. But the ball getting 10 feet further down the track in one second is precisely what we mean when we say the ball moves at 10 fps down the track. Who could doubt this?

Indeed, I encourage you not to doubt it until you have mastered the kinds of puzzles presented at the end of the preceding Chapter. Nevertheless, I call your attention to a dangerous phrase: "in one second". We have implicitly assumed that "in one second" means the same thing in the train frame as it does in the track frame. "Well," you will say, "of course it does. A second's a second." But suppose that's not true. Suppose "in a second" in the train frame means something different from "in a second" in the track frame. What happens to the argument we just gave? We would have to replace "in a second" by something like "in a second according to train-time" or "in a second according to track-time". The argument starts off fine, just a little more cumbersome:

> "If the ball moves down the train at 5 fps then in one second according to train time it gets 5 feet further down the train. And if the train moves down the track at 5 fps then in one second according to track time it gets 5 feet further down the track."

But then we come to:

> "So *in one second* the ball gets 10 feet further down the track — the 5 it gains on the train and the additional 5 the train gains on the track."

What can that italicized "in one second" mean here? The first 5 feet are gained in one second according to train time, the second 5 feet are gained in one second according to track time. Collapsing both into a single, unqualified "in one second" makes no sense. And indeed, when we get to the conclusion,

---

[6] I shall stop adding the cumbersome phrase "in the train frame" with the understanding then when we talk about the speed of the ball "down the train' we mean its speed in the train frame.

[7] Similarly, by the speed of the train "down the track" we mean its speed in the track frame.

> "But the ball getting 10 feet further down the track in one second is precisely what we mean when we say the ball moves at 10 fps down the track."

we see that this only works if the italicized "in one second" means in one second according to track time, since what we precisely mean when we say the ball moves at 10 fps down the track is that it moves 10 feet in one second according to track time. So the conclusion rests on being able to replace "in one second according to train time" by "in one second according to track time".

For the moment we will not pursue this any further. But please be aware that the simple rule (2.1) telling us how velocities combine is based on the implicit assumption that there is nothing problematic about the idea of a single unique notion of time that can be used equally well in any frame of reference. It was Einstein's great insight in 1905, that this apparently obvious assumption is, in fact, false. It's failure, however, is so slight as to be of no importance when all speeds concerned are small compared with that of light.

Last revised 1/14/01

3. The Speed of Light

When you turn on a light, how long does it take the light to get from the bulb to the things it illuminates? Galileo apparently tried to answer this by stationing two people with lanterns on top of distant mountains, a known distance d apart. Alice opens her lantern, Bob opens his the instant he sees Alice's, and Alice notes the time t that passes between the moment she opens hers and the moment she sees Bob's. To get the speed c with which the light moves from her mountain top to Bob's and back again, Alice just divides twice the distance between the mountains by the delay time t to get

$$c = 2d/t. \tag{3.1}$$

I don't know if Galileo worried about it, but there is a problem here: how much of the delay is due to the time it took the light to get from Alice to Bob and back, and how much is due to the speed of Bob's response—i.e. the time it takes the reception of Alice's light at Bob's eyes to reach his brain and be converted into a signal that reaches the muscles in his arms that operate the tendons that cause his fingers to open the shutter of his lantern.

There is an easy (but inspired) way to take care of this problem. Simply do the experiment again with Bob on a second mountain farther away from Alice. Bob's response time won't change (assuming the light from Alice is not now too dim to see clearly) so the increase in the delay time is entirely a result of the increase in the time it takes the light from the two lanterns to travel between the two more distance mountains.[1] Since this increase is just twice the increase in distance divided by the speed of light, Alice is back in a position to figure out the speed of light without having to know anything about Bob's response time. She simply uses (1) above with d being the *increase* in the distance between her and Bob in the two cases, and t being the *increase* in the time between her sending and receiving light signals.

But unfortunately, if she does this, Alicey will observe no discernible change in the delay time. Either it takes no time at all for light to travel the extra distance (i.e. the speed of light is infinite) or Bob's sluggish response takes so very much longer than the light travel time that Alice simply can't tell the difference between the two cases. And indeed, light travels so quickly and human response times are slow enough that two terrestrial mountains within view of each other are much too close for this method to work.

---

[1] I'm assuming here Alice's mountain and the two used by Bob are all on a single straight line. If not Alice has to do a slightly more complicated calculation.

Three centuries later Galileo's unsuccessful attempt was realized by replacing the two mountains by the earth and the moon. The moon is so far away that it takes radar[2] over two seconds to get there and bounce back. But by then the speed of light was known to high precision by other methods.

To measure the speed of light either you have to allow it to travel a great distance increase or you have to be able to make very accurate measurements of tiny intervals of time. The very first successful estimate of the speed of light came from using astronomical distances. Careful observations of the regularly occurring eclipses of the moons of Jupiter[3] (coincidentally discovered by Galileo) revealed that sometimes they lagged behind schedule by about ten minutes, and sometimes they came in ten minutes ahead. It was noted that they were ahead of schedule when the earth was closest to Jupiter, and behind when the earth was furthest away. So the time it takes light to cross the orbit of the Earth must be something like 20 minutes. This gives an estimate of some hundreds of thousands of miles per second for the speed of light. (Romer, 1676.)

In the 19th century a terrestrial measurement was done (Fizeau) by sharpening up the precision with which tiny time intervals could be measured. Imagine an axle with identical cog wheels at each end. Turn one of the wheels a little bit so that its teeth come exactly in between the gaps in the teeth of the other wheel. Because of this misalignment, if you try to send a thin beam of light parallel to the axle through a gap between the teeth of one wheel, it will be blocked by a tooth of the second wheel. But if you spin the whole thing extremely rapidly about the axle, you might hope that during the very tiny time it takes the light to pass between the two wheels, the second wheel will have turned just the tiny bit that is enough to allow the light that passed through the gap in the first wheel to get through a the gap in the second. After all, the wheels are spinning extremely fast and the teeth of the far away wheel have to move only a tiny fraction of a full turn to open up a passage for the light.

It turns out that for an axle short enough not to disturb the rather delicate alignment by a little twisting or bending, the light still travels too fast for this to work. However it is possible to introduce an enormous time-consuming detour for the light, in the form of a periscope-like perpendicular side journey with the help of four mirrors. When this was done, the sought for effect was observed, and the resulting estimate for the speed of light was in good agreement with that furnished by the earlier astronomical measurements.

Today we have highly sophisticated ways to measure the speed of light and know

[2] The speed of radar is the same as the speed of light. All electromagnetic radiation has the same speed in empty space.

[3] Moments when the moons disappeared behind Jupiter.

that it is 299,792,458 meters per second (m/s). Furthermore that is what it always shall be, because as of 1983 the meter has been *defined* to be not the disytance between two scratches on a platinum iridium bar carefully kept in a vacuum in Paris, but as the distance light travels in 1/299,792,458 seconds.[4]

There are some useful coincidences associated with the speed of light being 299,792,458 meters per second:

1. The number is comfortably close to 300 million m/s (unless you require a precision of better than 0.1%) or 300,000 kilometers per second (km/s). Physicists are very used to taking it to be 3×10^8 m/s. So much so that there is a legend that somebody once fouled up the report of a fine high precision experiment by using the number 3 rather than 2.9979 in converting the result into a more convenient form.

2. The corresponding English unit is about 186,000 miles per second. Since there are 5280 feet in a mile, there is good news for those in Washington and a few remote outposts elsewhere in the world who still resist the metric system. For this works out to about 982,000,000 feet per second. Thus within 2% accuracy the speed of light is 1 billion feet per second or, in more practical units, 1 foot per nanosecond.[5] A speed of 1 ft per nanosecond is actually relevant in setting limits to the size a computer can have if you want it to be really fast. Arithmetic operations are now being done in substantially less than a microsecond (a millionth of a second), nanosecond computers are just around the corner, and if you want to communicate with some remote part of the computer what you have just done before you do the next thing it had better not be more than half a foot away, since (as we shall see) no information can be transmitted faster than the speed of light.

In discussing issues related to speeds it is very useful to use units in which the speed of light assumes an especially simple form. In 1959 the foot was officially redefined to be exactly 0.3048 of a meter. Since, the speed of light is exactly 299,792,458 meters per

[4] The second is defined as the time it takes the light emitted by a certain atom—I forget which—under a particular set of circumstances—I forget what—to undergo a certain number of vibrations—I forget how many. The substantive point is that our unit of length (the meter) now is tied to our unit of time (the second). You might think that since the speed of light is now fixed forever by the definition of the meter, this means that there is no longer any point in striving to measure it more and more accurately. But such improved experiments now provide more and more accurate measurements of the length of a meter — better and better standards of length. The experiments remain just as important as they used to be. What has changed is the manner in which we describe what we have learned from them.

[5] A nanosecond (ns) is a billionth of a second.

second, if only people in 1959 had defined the foot to be 0.299792458 of a meter, a mere 1.64% shorter, then the speed of light would now be *exactly* one foot per nanosecond (1 f/n).[6] This unit will prove to be so useful for concrete examples, that for the purposes of Physics 209 *I hereby redefine the foot:*

Henceforth by one foot we shall mean the distance light travels in a nanosecond. A foot, if you will, is a light nanosecond (and a nanosecond, even more nicely, can be viewed as a light foot.) We shall revert to the clumsier term "light nanosecond" (or "P209 foot") if it ever becomes necessary to distinguish between our foot, and the conventional slightly larger foot, but I doubt that it will. If it deeply offends you to redefine the foot (as it did one referee of a paper I sent to the American Journal of Physics a few years ago) then you may define 0.299792458 meters (about 11.8 inches) to be one phoot, and think "phoot" whenever I say or write "foot".

There is something peculiar, and, as we shall see, extraordinary and remarkable about the unqualified assertion that the speed of light in empty space is 299,792,458 meters per second. Ordinarily when you specify a speed to such high precision and indeed when you mention any speed at all, the question "with respect to what" comes irresistibly to mind. After all the speed of an object depends on the frame of reference in which that speed is measured. A ball somebody throws while riding on a uniformly moving train has one speed with respect to the train, but quite another speed with respect to the tracks. In the case of light there are two obvious possible answers to the question "with respect to what?":

First obvious answer: The speed of light is 299,792,458 m/sec with respect to the source of the light. When you turn on a flashlight, the light it produces has a speed of 299,792,458 m/sec with respect to that flashlight. What else could it be? In much the same way, when one specifies the speed of a bullet, one always has in mind its speed with respect to the gun from which it has emerged.

Unfortunately this answer is contradicted by our current understanding of the electromagnetic character of light. In the 19th century there was a great unification of the laws of electricity and magnetism, completed by the Scottish physicist James Clerk Maxwell. Maxwell's equations led to the prediction that when electrically charged particles jiggled back and forth (as they do, for example inside a hot wire) they would emit radiant energy that travelled at a speed of about 300,000 kilometers per second. Since this speed was

[6] Sometimes more conveniently expressed as 1000 feet per microsecond (1000 f/μs). For comparison note that the speed of sound in ordinary air is about 1000 f/s. Light travels about a million times faster than sound.

numerically indistinguishable from the speed of light, it was natural to identify light with a particular form of such radiation (associated with a very rapid jiggling — almost a million billion times a second). Maxwell's equations quite unambiguously implied that *this speed was quite independent of the speed of the source of the radiation.* The speed of the light did not depend on whether the chunk of matter in which the charged particles were jiggling was stationary or moving toward or away from the direction in which the light was emitted.[7]

Second obvious answer: With respect to a light medium (historically called the ether).[8] The analogy here would be not to bullets from a gun, but to sound, which is a wave in the air. Like the speed of light, the speed of sound does not depend on the speed of the source of the sound. Sound moves at a definite speed with respect to the air, whose vibrations constitute and transmit that sound. If light is a vibration of something called the ether, then the speed of light should be with respect to that ether.

Since the Earth moves about the sun at a brisk clip of 30 km/sec in various directions, and the sun moves briskly about the center of our galaxy, it would be a remarkable coincidence if the earth just happened to be stationary in the rest frame of the ether. One would expect there to be a kind of "ether wind" blowing past the earth, leading to a dependence of the speed of light on earth on the direction of that wind.[9] Efforts to detect such a difference failed to yield a clearcut result, most famously in the the Michelson-Morley experiment of 1887. The measurements demonstrated that if the speed of light was fixed with respect to an ether, then at the time the experiment was performed, the earth, in spite of its complicated motion with respect to the galaxy, was improbably close to being at rest in the rest frame of that ether.[10]

---

[7] People had also noted that the regularity of certain astronomical motions as observed on earth was quite unaffected by whether the source of the light that enabled us to observe them was moving towards or aways from us, as it would be if the light travelled to us more slowly when the source was moving away than when the source was moving towards us.

[8] I digress to remark that 299,792,458 meters per second is the speed of light *in vacuum*. Light goes significantly slower in transparent media like water or glass, and even a little bit slower in air. Therefore this ether would have to be a sort of irreducible residue of otherwise empty space—what remains after you've removed everything it is possible to remove.

[9] Thus the speed of light on earth into the direction from which the ether wind was blowing ought to be a bit less than its speed along the direction of the wind.

[10] Stubborn people considered the possibility that the earth dragged the ether in its neighborhood along with it. But if that were so then the apparent positions of stars in the sky should shift through the year depending on the way in which the ether was being dragged by the Earth. No such shift is observed.

The importance of the Michelson-Morley experiment in the historical development of relativity has been hotly debated. In his famous 1905 paper setting forth relativity[11] Einstein alludes to it only once and then only in passing: "Examples of this sort, together with *unsuccessful attempts to determine any motion of the earth relative to the 'light medium',* lead to the conjecture that. . . ." (My italics.) The reference is hardly more than parenthetical. Such attempts have to be mentioned, because if they had been successful and unambiguously revealed a significant direction dependence to the velocity of light on earth, reflecting its motion through the ether, the theory of relativity would have been dead on arrival.

The "examples of this sort" that Einstein offered as the real motivation for his reexamination of the nature of time, were examples of the fact that the electric and magnetic behavior of matter does seem to be consistent with the principle of relativity, in spite of the then widespread view that there actually was a preferred inertial frame of reference for electromagnetic phenomena — namely the one in which the ether was stationary. The equations of electromagnetic theory were thought by many to be valid in that frame of reference and no other. Einstein noted, in effect, that even granting that assumption, a broad range of electromagnetic phenomena seemed to play out in exactly the same way in frames of reference other than the frame in which the ether was stationary. This led him to postulate that the laws of electromagnetism were, in fact, rigorously valid in arbitrary inertial frames of reference.[12] If this postulate were valid then, Einstein noted, "the introduction of a 'luminiferous ether' will prove to be superfluous" because there would be no way of determining the rest frame of the ether by any physical experiment involving electromagnetic phenomena.

But if Maxwell's equations are valid in any inertial frame of reference, and if they predict that electromagnetic radiation and light in particular propagate at a fixed speed that is independent of the speed of the source of the light, then light must propagate at the same speed in any inertial frame of reference. The answer to the question "with respect to what?" is, as we now know, "with respect to any inertial frame you like." The speed of light in vacuum is simply 299,792,458 m/sec in any inertial frame of reference, regardless of how fast the source of the light is moving, and regardless of the choice of frame of reference

[11] "Zur Electrodynamik Bewegter Körper" ("On the Electrodynamics of Moving Bodies"), Annalen der Physik **17**, 132-148.

[12] It is this specific postulate — that what we now call the principle of relativity applies to electromagnetism as well as to Newtonian mechanics (where everybody agreed that it was indeed valid) — that Einstein named the "Principle of Relativity" (*Prinzip der Relativität*).

in which the measurement of the speed of the light is made. If, for example, you race after the light in a rocket at 10 km/sec you do not reduce its speed away from you to 299,782 km/sec. It still recedes from you at 299,792 km/sec.[13]

How can this be? How can there be a speed[14] c with the property that if something moves with speed c then it must have the speed c in any inertial frame of reference? This is highly counterintuitive. Indeed, "counterintuitive" is too weak a word. It seems downright impossible. One of the central aims of our study of relativity will be to remove this sense of impossibility, and see how it can, in fact, make perfect sense.

To do this we must look very closely and critically at what it actually means to "have a speed" with respect to a particular frame of reference. When we say that an object moves uniformly with a certain speed s, we mean that it goes a certain distance D in a certain time T and that the distance and time are related by $D/T = s$. We are therefore led, inexorably, to examine carefully how one actually measures such distances and how one actually measures such times.

Let P be a valid procedure for carrying out the time and distance measurements that allow one to determine the speed of an object in a given inertial frame. Let Bob, carrying out the procedure P in the frame of reference of a space station, measure the speed of a pulse of light as it zooms off into space. He will find that it moves at about 299,792 km/s. Suppose Alice flies swiftly after the light at a speed Bob determines to be 792 km/s. Bob will then (correctly) note that in each second the light gets an additional 299,792 km away from him and Alice gets an additional 792 km away, so that the distance between Alice and the light is growing at only 299,000 km/sec. But if Alice carries out the same procedure P in the frame of reference of her rocket ship, she will find that the speed of the light is 299,792 km/s, so that the distance between her and the light is growing at the full 299,792 km/sec.

How are we to account for this discrepancy? Obviously the methods Alice uses to measure distances and times must be different from those used by Bob. But don't they both use exactly the same procedure P? Yes, but you have to think about what "exactly

---

[13] It is only the speed of light in *vacuum* that has this special property. The speed of light in water *does* depend on how fast you are moving through the water. Indeed, what is special is not light, but the speed $c = 299{,}792{,}458$ m/s. When one simply says "speed of light" without any qualification one almost always means the speed of light in vacuum, 299,792,458 m/s.

[14] Everybody calls the speed of light in vacuum c (as in, most famously, "$E = mc^2$", about which there will be more to say later on). I always thought c stood for "constant", reflecting the fact that it doesn't vary from one frame of reference to another. But perhaps it stands for *celeritas*—Latin for speed, as in "celerity" or "accelerate".

the same" means. If Bob, for example, uses clocks that are stationary in the frame of his space station to measure times, then if Alice uses *exactly* the same procedure she must use clocks that are stationary in the frame of *her* rocket ship. Thus in Bob's frame of reference Alice's clocks are moving, while his are not.[15] Similar considerations apply to the meter sticks they might use to measure distances. The not terribly subtle but easily overlooked point is that Bob's procedure *as described in Bob's frame of reference* must be exactly the same as Alice's procedure *as described in Alice's frame of reference.* But Alice's procedure *as described in Bob's frame of reference* is not exactly the same as Bob's procedure *as described in Bob's frame of reference.* It is this difference that makes it possible for either Bob or Alice to account, in an entirely rational way, for the discrepancy.

The fact that Alice and Bob, using different frames of reference, both find exactly the same speed for one and the same pulse of light appears paradoxical only if you assume several things, as everybody implicitly did until the year 1905, about the relation between the clocks and meter sticks used by Alice and Bob:

(1) The procedure Alice uses to synchronize all the clocks in her frame of reference gives a set of clocks that Bob agrees are synchronized when he tests them against a set of clocks that he has synchronized using the same[16] procedure in his own frame of reference.

(2) The rate of a clock, as determined in Bob's frame of reference is independent of how fast that clock moves with respect to Bob;

(3) The length of a meter stick as determined in Bob's frame of reference is independent of how fast that meter stick moves with respect to Bob,

If any of these assumptions is false, then we must reexamine the way in which the speed of an object changes as one changes the speed of the frame of reference in which the speed of that object is measured. We now know that *all three* of these assumptions are false. The special theory of relativity gives a quantitative specification of the way in which they fail, and how, when they are suitably corrected, one emerges with a simple and coherent picture of space and time measurements that is entirely in accord with the existence of an invariant speed — a speed that is the same in all inertial frames of reference.

The traditional (and simplest) way to arrive at this picture — the way we shall be taking and the way Einstein used — is simply to accept as a working hypothesis that in any inertial frame of reference, any procedure that correctly measures the speed of light in vacuum must give 299,792,458 m/s. We shall accept the strange fact that if Alice and Bob both measure the speed of the same pulse of light, they will both find it to be 299,792,458

[15] And, of course, vice-versa: in Alice's frame Bob's clocks are moving and hers are not.

[16] "Same" in that same tricky sense—that he does the same thing with respect to *his* frame of reference as Alice does with respect to *her* frame of reference.

m/s even though Alice and her measuring instruments may be moving at high speed with respect to Bob and his. By tentatively accepting this extremely peculiar fact, and insisting that the principle of relativity must nevertheless remain valid, we will be able to *deduce* the precise way in which each of the three assumptions about the behavior of moving clocks and meter sticks must be modified. Once this is done, and the corrected versions of these three assumptions are understood, the strange fact will cease to appear strange. This will be our preoccupation for the next four or five weeks.

This remarkable property of light — that its speed does not depend on the frame of reference in which it is measured — is today called the "Principle of the Constancy of the Velocity of Light". The special theory of relativity is said to rest on two principles: the Principle of Relativity and the Principle of the Constancy of the Velocity of Light. In Einstein's great 1905 paper he did not use the word *Prinzip* for this second principle (as he did for the first). He characterized each principle as a "postulate" (*Voraussetzung*). His second postulate was that light in empty space moves with a velocity that is independent of the velocity of the body that emitted the light. This is tantamount to the second "Principle" when it is conjoined with the first, which Einstein stated as the postulate that the concept of absolute rest has no more meaning for electromagnetic phenomena than it does for phenomena in ordinary Newtonian mechanics.[17]

[17] If you know a little German and are interested in seeing how Einstein sets up the problem, you can download the text of his famous paper (*Annalen der Physik und Chemie, IV. Folge, Band 17* (1905) 891-921) from

http://www.wiley-vch.de/berlin/journals/adp/historic.html.

Everything I've quoted is from the first page and a half.