Notes on Special Relativity

Physics 252, Michael Fowler, University of Virginia

March 21, 2008

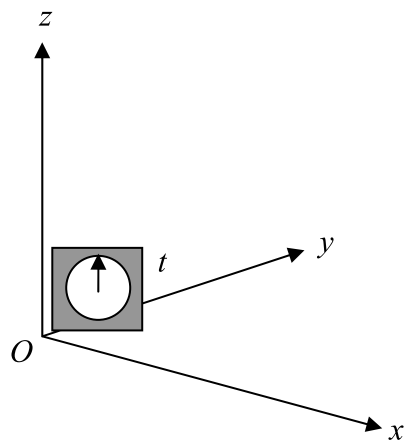
Frames of Reference and Newton's Laws

The cornerstone of the theory of special relativity is the *Principle of Relativity*:

The Laws of Physics are the same in all inertial frames of reference.

We shall see that many surprising consequences follow from this innocuous looking statement.

Let us first, however, briefly review Newton's mechanics in terms of frames of reference.



A frame of reference

A "frame of reference" is just a set of coordinates: something you use to measure the things that matter in Newtonian problems, that is to say, positions and velocities, so we also need a clock.

A point in space is specified by its three coordinates (x, y, z) and an "event" like, say, a little explosion, by a place and time: (x, y, z, t).

An inertial frame is defined as one in which Newton's law of inertia holds—that is, any body which isn't being acted on by an outside force stays at rest if it is initially at rest, or continues to move at a constant velocity if that's what it was doing to begin with. An example of a *non*-inertial frame is a rotating frame, such as a carousel.

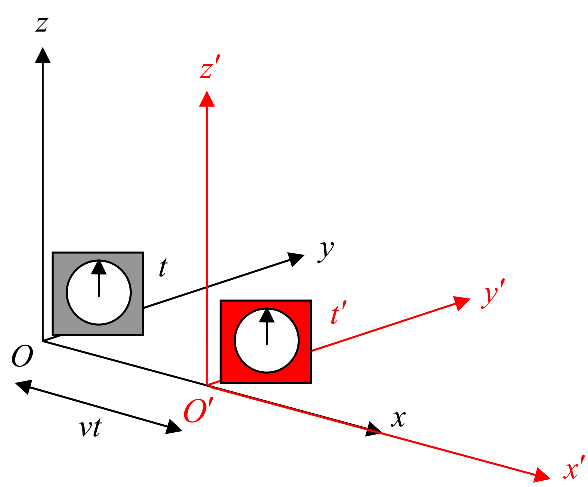
The "laws of physics" we shall consider first are those of Newtonian mechanics, as expressed by Newton's Laws of Motion, with gravitational forces and also contact forces from objects pushing against each other. For example, knowing the universal gravitational constant from experiment (and the masses involved), it is possible from Newton's Second Law,

$$force = mass \times acceleration,$$

to predict future planetary motions with great accuracy.

Suppose we know from experiment that these laws of mechanics are true in one frame of reference. How do they look in another frame, moving with respect to the first frame? To find out, we have to figure out how to get from position, velocity and acceleration in one frame to the corresponding quantities in the second frame.

Obviously, the two frames must have a constant relative velocity, otherwise the law of inertia won't hold in both of them. Let's choose the coordinates so that this velocity is along the x-axis of both of them.



Two frames of reference relatively displaced along the *x*-axis

Notice we also throw in a clock with each frame.

Suppose S' is proceeding relative to S at speed v along the x-axis. For convenience, let us label the moment when O' passes O as the zero point of timekeeping.

Now what are the coordinates of the event (x, y, z, t) in S'? It's easy to see $t' = t$—we synchronized the clocks when O' passed O. Also, evidently, $y' = y$ and $z' = z$, from the figure. We can also see that $x = x' + vt$. Thus (x, y, z, t) in S corresponds to (x', y', z', t') in S', where

$$
\begin{aligned}
x' &= x - vt \\
y' &= y \\
z' &= z \\
t' &= t.
\end{aligned}
$$

That's how *positions* transform; these are known as the *Galilean* transformations.

What about *velocities* ? The velocity in S' in the x' direction

$$u'_x = \frac{dx'}{dt'} = \frac{dx'}{dt} = \frac{d}{dt}(x - vt) = \frac{dx}{dt} - v = u_x - v.$$

This is obvious anyway: it's just the addition of velocities formula

$$u_x = u'_x + v.$$

How does *acceleration* transform?

$$\frac{du'_x}{dt'} = \frac{du'_x}{dt} = \frac{d}{dt}(u_x - v) = \frac{du_x}{dt}$$

since v is constant.

That is to say,

$$a'_x = a_x$$

the *acceleration is the same* in both frames. This again is obvious—the acceleration is the rate of change of velocity, and the velocities of the same particle measured in the two frames differ by a *constant* factor-the relative velocity of the two frames.

If we now look at the motion under gravitational forces, for example,

$$m_1\vec{a} = \frac{Gm_1m_2}{r^2}\hat{\vec{r}}$$

we get the same law on going to another inertial frame because *every term in the above equation stays the same*.

Note that $m\vec{a}$ is the rate of change of momentum—this is the same in both frames. So, in a collision, say, if total momentum is conserved in one frame (the sum of individual rates of change of momentum is zero) the same is true in *all* inertial frames.

The Speed of Light

Early Ideas about Light Propagation

As we shall soon see, attempts to measure the speed of light played an important part in the development of the theory of special relativity, and, indeed, the speed of light is central to the theory.

The first recorded discussion of the speed of light (I think) is in Aristotle, where he quotes Empedocles as saying the light from the sun must take some time to reach the earth, but Aristotle himself apparently disagrees, and even Descartes thought that light traveled instantaneously. Galileo, unfairly as usual, in *Two New Sciences* (page 42) has Simplicio stating the Aristotelian position,

SIMP. Everyday experience shows that the propagation of light is instantaneous; for when we see a piece of artillery fired at great distance, the flash reaches our eyes without lapse of time; but the sound reaches the ear only after a noticeable interval.

Of course, Galileo points out that in fact nothing about the speed of light can be deduced from this observation, except that light moves faster than sound. He then goes on to suggest a possible way to measure the speed of light. The idea is to have two people far away from each other, with covered lanterns. One uncovers his lantern, then the other immediately uncovers his on seeing the light from the first. This routine is to be practiced with the two close together, so they will get used to the reaction times involved, then they are to do it two or three miles apart, or even further using telescopes, to see if the time interval is perceptibly lengthened. Galileo claims he actually tried the experiment at distances less than a mile, and couldn't detect a time lag. From this one can certainly deduce that light travels at least ten times faster than sound.

Measuring the Speed of Light with Jupiter's Moons

The first real measurement of the speed of light came about half a century later, in 1676, by a Danish astronomer, Ole Römer, working at the Paris Observatory. He had made a systematic study of Io, one of the moons of Jupiter, which was eclipsed by Jupiter at regular intervals, as Io went around Jupiter in a circular orbit at a steady rate. Actually, Römer found, for several months the eclipses lagged more and more behind the expected time, until they were running about eight minutes late, then they began to pick up again, and in fact after about six months were running eight minutes early. The cycle then repeated itself. Römer realized the significance of the time involved-just over one year. This time period had nothing to do with Io, but was the time between successive closest approaches of earth in its orbit to Jupiter. The eclipses were furthest behind the predicted times when the earth was furthest from Jupiter.

The natural explanation was that the light from Io (actually reflected sunlight, of course) took time to reach the earth, and took the longest time when the earth was furthest away. From his observations, Römer concluded that light took about twenty-two minutes to cross the earth's orbit. This was something of an overestimate, and a few years later Newton wrote in the *Principia* (Book I, section XIV): "For it is now certain from the phenomena of Jupiter's satellites, confirmed by the observations of different astronomers, that light is propagated in succession (NOTE: I think this means at finite speed) and requires about seven or eight minutes to travel from the sun to the earth." This is essentially the correct value.

Of course, to find the speed of light it was also necessary to know the distance from the earth to the sun. During the 1670's, attempts were made to measure the parallax of Mars, that is, how far it shifted against the background of distant stars when viewed simultaneously from two different places on earth at the same time. This (very slight) shift could be used to find the distance of Mars from earth, and hence the distance to the sun, since all *relative* distances in the solar system had been established by observation and geometrical analysis. According to Crowe (*Modern Theories of the Universe*, Dover, 1994, page 30), they concluded that the distance to the sun was between 40 and 90 million miles. Measurements presumably converged on the correct value of about 93 million miles soon after that, because it appears Römer (or perhaps Huygens, using Römer's data a short time later) used the correct value for the distance, since the speed of light was calculated to be 125,000 miles per second, about three-quarters of the correct value of 186,300 miles per second. This error is fully accounted for by taking the time light needs to cross the earth's orbit to be twenty-two minutes (as Römer did) instead of the correct value of sixteen minutes.

Starlight and Rain

The next substantial improvement in measuring the speed of light took place in 1728, in England. An astronomer James Bradley, sailing on the Thames with some friends, noticed that the little pennant on top of the mast changed position each time the boat put about, even though the wind was steady. He thought of the boat as the earth in orbit, the wind as starlight coming from some distant star, and reasoned that the apparent direction the starlight was "blowing" in would depend on the way the earth was moving. Another possible analogy is to imagine the starlight as a steady downpour of rain on a windless day, and to think of yourself as walking around a circular path at a steady pace. The apparent direction of the incoming rain will not be vertically downwards-more will hit your front than your back. In fact, if the rain is falling at, say, 15 mph, and you are walking at 3 mph, to you as observer the rain will be coming down at a slant so that it has a vertical speed of 15 mph, and a horizontal speed towards you of 3 mph. Whether it is slanting down from the north or east or whatever at any given time depends on where you are on the circular path at that moment. Bradley reasoned that the apparent direction of incoming starlight must vary in just this way, but the angular change would be a lot less dramatic. The earth's speed in orbit is about 18 miles per second, he knew from Römer's work that light went at about 10,000 times that speed. That meant that the angular variation in apparent incoming direction of starlight was about the magnitude of the small angle in a right-angled triangle with one side 10,000 times longer than the other, about one two-hundredth of a degree. Notice this would have been just at the limits of Tycho's measurements, but the advent of the telescope, and general improvements in engineering, meant this small angle was quite accurately measurable by Bradley's time, and he found the velocity of light to be 185,000 miles per second, with an accuracy of about one percent.

Fast Flickering Lanterns

The problem is, all these astronomical techniques do not have the appeal of Galileo's idea of two guys with lanterns. It would be reassuring to measure the speed of a beam of light between two points on the ground, rather than making somewhat indirect deductions based on apparent slight variations in the positions of stars. We can see, though, that if the two lanterns are ten miles apart, the time lag is of order one-ten thousandth of a second, and it is difficult to see how to arrange that. This technical problem was solved in France about 1850 by two rivals, Fizeau and Foucault, using slightly different techniques. In Fizeau's apparatus, a beam of light shone between the teeth of a rapidly rotating toothed wheel, so the "lantern" was constantly being covered and uncovered. Instead of a second lantern far away, Fizeau simply had a mirror, reflecting the beam back, where it passed a second time between the teeth of the wheel. The idea was, the blip of light that went out through one gap between teeth would only make it back through the same gap if the teeth had not had time to move over significantly during the round trip time to the far away mirror. It was not difficult to make a wheel with a hundred teeth, and to rotate it hundreds of times a second, so the time for a tooth to move over could be arranged to be a fraction of one ten thousandth of a second. The method worked. Foucault's method was based on the same general idea, but instead of a toothed wheel, he shone the beam on to a rotating mirror. At one point in the mirror's rotation, the reflected beam fell on a distant mirror, which reflected it right back to the rotating mirror, which meanwhile had turned through a small angle. After this second reflection from the rotating mirror, the position of the beam was carefully measured. This made it possible to figure out how far the mirror had turned during the time it took the light to make the round trip to the distant mirror, and since the rate of rotation of the mirror was known, the speed of light could be figured out. These techniques gave the speed of light with an accuracy of about 1,000 miles per second.

Albert Abraham Michelson

Albert Michelson was born in 1852 in Strzelno, Poland. His father Samuel was a Jewish merchant, not a very safe thing to be at the time. Purges of Jews were frequent in the neighboring towns and villages. They decided to leave town. Albert's fourth birthday was celebrated in Murphy's Camp, Calaveras County, about fifty miles south east of Sacramento, a place where five million dollars worth of gold dust was taken from one four acre lot. Samuel prospered selling supplies to the miners. When the gold ran out, the Michelsons moved to Virginia City, Nevada, on the Comstock lode, a silver mining town. Albert went to high school in San Francisco. In 1869, his father spotted an announcement in the local paper that Congressman Fitch would be appointing a candidate to the Naval Academy in Annapolis, and inviting applications. Albert applied but did not get the appointment, which went instead to the son of a civil war veteran. However, Albert knew that President Grant would also be appointing ten candidates himself, so he went east on the just opened continental railroad to try his luck. Unknown to Michelson, Congressman Fitch wrote directly to Grant on his behalf, saying this would really help get the Nevada Jews into the Republican party. This argument proved persuasive. In fact, by the time Michelson met with Grant, all ten scholarships had been awarded, but the President somehow came up with another one. Of the incoming class of ninety-two, four years later twenty-nine graduated. Michelson placed first in optics, but twenty-fifth in seamanship. The Superintendent of the Academy, Rear Admiral Worden, who had commanded the Monitor in its victory over the Merrimac, told Michelson: “If in the future you’d give less attention to those scientific things and more to your naval gunnery, there might come a time when you would know enough to be of some service to your country.”

Sailing the Silent Seas: Galilean Relativity

Shortly after graduation, Michelson was ordered aboard the USS *Monongahela*, a sailing ship, for a voyage through the Caribbean and down to Rio. According to the biography of Michelson written by his daughter (*The Master of Light*, by Dorothy Michelson Livingston, Chicago, 1973) he thought a lot as the ship glided across the quiet Caribbean about whether one could decide in a closed room inside the ship whether or not the vessel was moving. In fact, his daughter quotes a famous passage from Galileo on just this point:

[SALV.] Shut yourself up with some friend in the largest room below decks of some large ship and there procure gnats, flies, and other such small winged creatures. Also get a great tub full of water and within it put certain fishes; let also a certain bottle be hung up, which drop by drop lets forth its water into another narrow-necked bottle placed underneath. Then, the ship lying still, observe how those small winged animals fly with like velocity towards all parts of the room; how the fish swim indifferently towards all sides; and how the distilling drops all fall into the bottle placed underneath. And casting anything toward your friend, you need not throw it with more force one way than another, provided the distances be equal; and leaping with your legs together, you will reach as far one way as another. Having observed all these particulars, though no man doubts that, so long as the vessel stands still, they ought to take place in this manner, make the ship move with what velocity you please, so long as the motion is uniform and not fluctuating this way and that. You will not be able to discern the least alteration in all the forenamed effects, nor can you gather by any of them whether the ship moves or stands still. ...in throwing something to your friend you do not need to throw harder if he is towards the front of the ship from you... the drops from the upper bottle still fall into the lower bottle even though the ship may have moved many feet while the drop is in the air ... Of this correspondence of effects the cause is that the ship’s motion is common to all the things contained in it and to the air also; I mean if those things be shut up in the room; but in case those things were above the deck in the open air, and not obliged to follow the course of the ship, differences would be observed, ... smoke would stay behind... .

[SAGR.] Though it did not occur to me to try any of this out when I was at sea, I am sure you are right. I remember being in my cabin wondering a hundred times whether the ship was moving or not, and sometimes I imagined it to be moving one way when in fact it was moving the other way. I am therefore satisfied that no experiment that can be done in a closed cabin can determine the speed or direction of motion of a ship in steady motion.

I have paraphrased this last remark somewhat to clarify it. This conclusion of Galileo's, that everything looks the same in a closed room moving at a steady speed as it does in a closed room at rest, is called *The Principle of Galilean Relativity*. We shall be coming back to it.

Michelson Measures the Speed of Light

On returning to Annapolis from the cruise, Michelson was commissioned Ensign, and in 1875 became an instructor in physics and chemistry at the Naval Academy, under Lieutenant Commander William Sampson. Michelson met Mrs. Sampson's niece, Margaret Heminway, daughter of a very successful Wall Street tycoon, who had built himself a granite castle in New Rochelle, NY. Michelson married Margaret in an Episcopal service in New Rochelle in 1877.

At work, lecture demonstrations had just been introduced at Annapolis. Sampson suggested that it would be a good demonstration to measure the speed of light by Foucault's method. Michelson soon realized, on putting together the apparatus, that he could redesign it for much greater accuracy, but that would need money well beyond that available in the teaching demonstration budget. He went and talked with his father in law, who agreed to put up $2,000. Instead of Foucault's 60 feet to the far mirror, Michelson had about 2,000 feet along the bank of the Severn, a distance he measured to one tenth of an inch. He invested in very high quality lenses and mirrors to focus and reflect the beam. His final result was 186,355 miles per second, with possible error of 30 miles per second or so. This was twenty times more accurate than Foucault, made the New York Times, and Michelson was famous while still in his twenties. In fact, this was accepted as the most accurate measurement of the speed of light for the next forty years, at which point Michelson measured it again.

The Michelson-Morley Experiment

The Nature of Light

As a result of Michelson's efforts in 1879, the speed of light was known to be 186,350 miles per second with a likely error of around 30 miles per second. This measurement, made by timing a flash of light travelling between mirrors in Annapolis, agreed well with less direct measurements based on astronomical observations. Still, this did not really clarify the *nature* of light. Two hundred years earlier, Newton had suggested that light consists of tiny *particles* generated in a hot object, which spray out at very high speed, bounce off other objects, and are detected by our eyes. Newton's arch-enemy Robert Hooke, on the other hand, thought that light must be a kind of *wave motion*, like sound. To appreciate his point of view, let us briefly review the nature of sound.

The Wavelike Nature of Sound

Actually, *sound* was already quite well understood by the ancient Greeks. The essential point they had realized is that sound is generated by a vibrating material object, such as a bell, a string or a drumhead. Their explanation was that the vibrating drumhead, for example, alternately pushes and pulls on the air directly above it, sending out waves of compression and decompression (known as rarefaction), like the expanding circles of ripples from a disturbance on the surface of a pond. On reaching the ear, these waves push and pull on the eardrum with the same frequency (that is to say, the same number of pushes per second) as the original source was vibrating at, and nerves transmit from the ear to the brain both the intensity (loudness) and frequency (pitch) of the sound.

There are a couple of special properties of sound waves (actually any waves) worth mentioning at this point. The first is called *interference*. This is most simply demonstrated with water waves. If you put two fingers in a tub of water, just touching the surface a foot or so apart, and vibrate them at the same rate to get two expanding circles of ripples, you will notice that where the ripples overlap there are quite complicated patterns of waves formed. The essential point is that at those places where the wave-crests from the two sources arrive at the same time, the waves will work together and the water will be very disturbed, but at points where the crest from one source arrives at the same time as the wave trough from the other source, the waves will cancel each other out, and the water will hardly move. You can hear this effect for sound waves by playing a constant note through stereo speakers. As you move around a room, you will hear quite large variations in the intensity of sound. Of course, reflections from walls complicate the pattern. This large variation in volume is *not* very noticeable when the stereo is playing music, because music is made up of many frequencies, and they change all the time. The different frequencies, or notes, have their quiet spots in the room in different places. The other point that should be mentioned is that high frequency tweeter-like sound is much more *directional* than low frequency woofer-like sound. It really doesn't matter where in the room you put a low-frequency woofer—the sound seems to be all around you anyway. On the other hand, it is quite difficult to get a speaker to spread the high notes in all directions. If you listen to a cheap speaker, the high notes are loudest if the speaker is pointing right at you. A lot of effort has gone into designing tweeters, which are small speakers especially designed to broadcast high notes over a wide angle of directions.

Is Light a Wave?

Bearing in mind the above minireview of the properties of waves, let us now reconsider the question of whether light consists of a stream of particles or is some kind of wave. The strongest argument for a particle picture is that light travels in straight lines. You can hear around a corner, at least to some extent, but you certainly can't see.

Furthermore, no wave-like interference effects are very evident for light. Finally, it was long known, as we have mentioned, that sound waves were compressional waves in air. If light is a wave, just what is waving? It clearly isn't just air, because light reaches us from the sun, and indeed from stars, and we know the air doesn't stretch that far, or the planets would long ago have been slowed down by air resistance.

Despite all these objections, it was established around 1800 that light *is* in fact some kind of wave. The reason this fact had gone undetected for so long was that the wavelength is *really* short, about one fifty-thousandth of an inch. In contrast, the shortest wavelength sound detectable by humans has a wavelength of about half an inch. The fact that light travels in straight lines is in accord with observations on sound that the higher the frequency (and shorter the wavelength) the greater the tendency to go in straight lines. Similarly, the interference patterns mentioned above for sound waves or ripples on a pond vary over distances of the same sort of size as the wavelengths involved. Patterns like that would not normally be noticeable for light because they would be on such a tiny scale. In fact, it turns out, there *are* ways to see interference effects with light. A familiar example is the many colors often visible in a soap bubble. These come about because looking at a soap bubble you see light reflected from both sides of a very thin film of water—a thickness that turns out to be comparable to the wavelength of light. The light reflected from the lower layer has to go a little further to reach your eye, so that light wave must wave an extra time or two before getting to your eye compared with the light reflected from the top layer. What you actually *see* is the *sum* of the light reflected from the top layer and that reflected from the bottom layer. Thinking of this now as the sum of two sets of waves, the light will be bright if the crests of the two waves arrive together, dim if the *crests* of waves reflected from the top layer arrive simultaneously with the *troughs* of waves reflected from the bottom layer. Which of these two possibilities actually occurs for reflection from a particular bit of the soap film depends on just how much further the light reflected from the lower surface has to travel to reach your eye compared with light from the upper surface, and that depends on the angle of reflection and the thickness of the film. Suppose now we shine *white* light on the bubble. White light is made up of all the colors of the rainbow, and these different colors have different wavelengths, so we see colors reflected, because for a particular film, at a particular angle, some colors will be reflected brightly (the crests will arrive together), some dimly, and we will see the ones that win.

If Light *is* a Wave, What is Waving?

Having established that light is a wave, though, we still haven't answered one of the major objections raised above. Just what is waving? We discussed sound waves as waves of compression in air. Actually, that is only one case—sound will also travel through liquids, like water, and solids, like a steel bar. It is found experimentally that, other things being equal, sound travels faster through a medium that is harder to compress: the material just springs back faster and the wave moves through more rapidly. For media of equal springiness, the sound goes faster through the less heavy medium, essentially because the same amount of springiness can push things along faster in a lighter material. So when a sound wave passes, the material—air, water or solid—waves as it goes through. Taking this as a hint, it was natural to suppose that light must be just waves in some mysterious material, which was called the *aether*, surrounding and permeating everything. This aether must also fill all of space, out to the stars, because we can see them, so the medium must be there to carry the light. (We could never *hear* an explosion on the moon, however loud, because there is no air to carry the sound to us.) Let us think a bit about what properties this aether must have. Since light travels so fast, it must be very light, and very hard to compress. Yet, as mentioned above, it must allow solid bodies to pass through it freely, without aether resistance, or the planets would be slowing down. Thus we can picture it as a kind of ghostly wind blowing through the earth. But how can we prove any of this? Can we detect it?

Detecting the Aether Wind: the Michelson-Morley Experiment

Detecting the aether wind was the next challenge Michelson set himself after his triumph in measuring the speed of light so accurately. Naturally, something that allows solid bodies to pass through it freely is a little hard to get a grip on. But Michelson realized that, just as the speed of sound is relative to the air, so the speed of light must be relative to the aether. This must mean, if you could measure the speed of light accurately enough, you could measure the speed of light travelling upwind, and compare it with the speed of light travelling downwind, and the difference of the two measurements should be twice the windspeed. Unfortunately, it wasn't that easy. All the recent accurate measurements had used light travelling to a distant mirror and coming back, so if there was an aether wind along the direction between the mirrors, it would have opposite effects on the two parts of the measurement, leaving a very small overall effect. There was no technically feasible way to do a one-way determination of the speed of light.

At this point, Michelson had a very clever idea for detecting the aether wind. As he explained to his children (according to his daughter), it was based on the following puzzle:

Suppose we have a river of width w (say, 100 feet), and two swimmers who both swim at the same speed v feet per second (say, 5 feet per second). The river is flowing at a steady rate, say 3 feet per second. The swimmers race in the following way: they both start at the same point on one bank. One swims directly across the river to the closest point on the opposite bank, then turns around and swims back. The other stays on one side of the river, swimming upstream a distance (measured along the bank) exactly equal to the width of the river, then swims back to the start. Who wins?

Let's consider first the swimmer going upstream and back. Going 100 feet upstream, the speed relative to the bank is only 2 feet per second, so that takes 50 seconds. Coming back, the speed is 8 feet per second, so it takes 12.5 seconds, for a total time of 62.5 seconds.

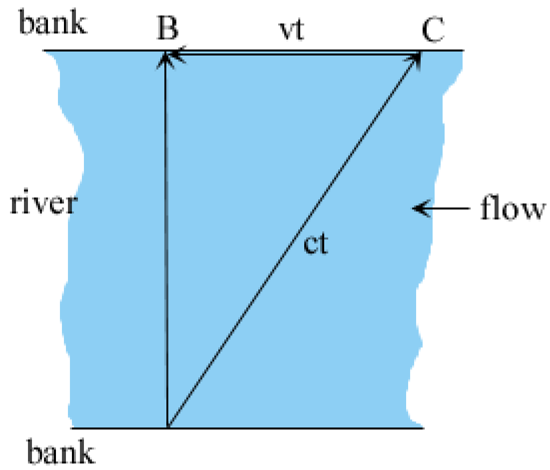


Figure 1: In time *t*, the swimmer has moved *ct* relative to the water, and been carried downstream a distance *vt*.

The swimmer going across the flow is trickier. It won't do simply to aim directly for the opposite bank-the flow will carry the swimmer downstream. To succeed in going directly across, the swimmer must actually aim upstream at the correct angle (of course, a real swimmer would do this automatically). Thus, the swimmer is going at 5 feet per second, at an angle, relative to the river, and being carried downstream at a rate of 3 feet per second. If the angle is correctly chosen so that the net movement is directly across, in one second the swimmer must have moved *four feet* across: the distances covered in one second will form a 3,4,5 triangle. So, at a crossing rate of 4 feet per second, the swimmer gets across in 25 seconds, and back in the same time, for a total time of 50 seconds. The cross-stream swimmer wins. This turns out to true whatever their swimming speed. (Of course, the race is only possible if they can swim faster than the current!)

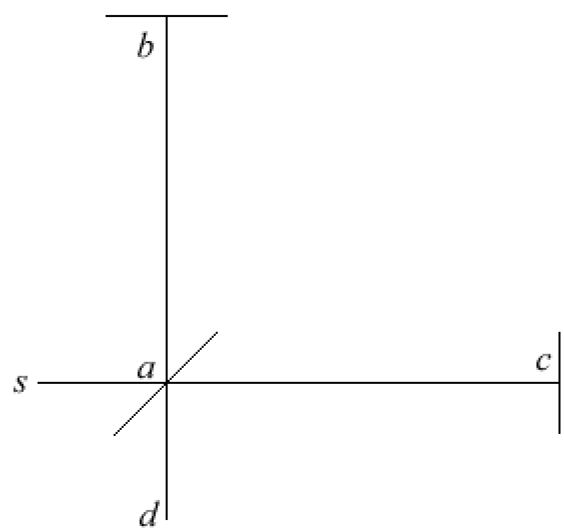


Figure 2: This diagram is from the original paper. The source of light is at *s*, the 45 degree line is the half-silvered mirror, *b* and *c* are mirrors and d the observer.

Michelson's great idea was to construct an exactly similar race for pulses of light, with the aether wind playing the part of the river. The scheme of the experiment is as follows: a pulse of light is directed at an angle of 45 degrees at a half-silvered, half transparent mirror, so that half the pulse goes on through the glass, half is reflected. These two half-pulses are the two swimmers. They both go on to distant mirrors which reflect them back to the half-silvered mirror. At this point, they are again half reflected and half transmitted, but a telescope is placed behind the half-silvered mirror as shown in the figure so that half of each half-pulse will arrive in this telescope. Now, if there is an aether wind blowing, someone looking through the telescope should see the halves of the two half-pulses to arrive at slightly different times, since one would have gone more upstream and back, one more across stream in general. To maximize the effect, the whole apparatus, including the distant mirrors, was placed on a large turntable so it could be swung around.

Let us think about what kind of time delay we expect to find between the arrival of the two half-pulses of light. Taking the speed of light to be c miles per second relative to the aether, and the aether to be flowing at v miles per second through the laboratory, to go a distance w miles upstream will take $w/(c-v)$ seconds, then to come back will take $w/(c+v)$ seconds. The total roundtrip time upstream and downstream is the sum of these, which works out to be $2wc/(c^2-v^2)$, which can also be written $(2w/c)\times 1/(1-v^2/c^2)$. Now, we can safely assume the speed of the aether is much less than the speed of light, otherwise it would have been noticed long ago, for example in timing of eclipses of Jupiter's satellites. This means v^2/c^2 is a very small number, and we can use some handy mathematical facts to make the algebra a bit easier. First, if x is very small compared to 1, $1/(1-x)$ is very close to $1+x$. (You can check it with your calculator.) Another fact we shall need in a minute is that for small x, the square root of $1+x$ is very close to $1+x/2$.

Putting all this together,

$$\text{upstream-downstream roundtrip time } \cong \frac{2w}{c}\times\left(1+\frac{v^2}{c^2}\right).$$

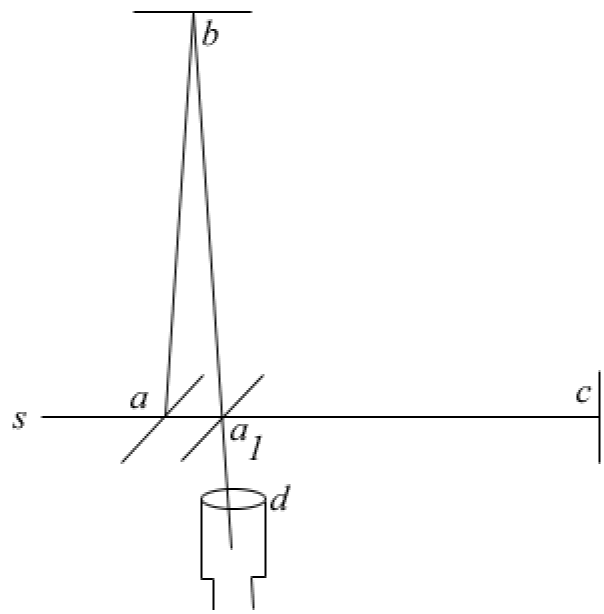


Figure 3 This is also from the original paper, and shows the expected path of light relative to the aether with an aether wind blowing.

Now, what about the cross-stream time? The actual cross-stream speed must be figured out as in the example above using a right-angled triangle, with the hypoteneuse equal to the speed *c*, the shortest side the aether flow speed *v*, and the other side the cross-stream speed we need to find the time to get across. From Pythagoras' theorem, then, the cross-stream speed is the square root of (c^2-v^2).

Since this will be the same both ways, the roundtrip cross-stream time will be

$$2w/\sqrt{c^2-v^2}.$$

This can be written in the form

$$\frac{2w}{c}\frac{1}{\sqrt{1-v^2/c^2}} \cong \frac{2w}{c}\frac{1}{1-\left(v^2/2c^2\right)} \cong \frac{2w}{c}\left(1+\frac{v^2}{2c^2}\right)$$

where the two successive approximations, valid for $v/c = x \ll 1$, are $\sqrt{1-x} \cong 1-\left(x/2\right)$ and $1/\left(1-x\right) \cong 1+x$.

Therefore the

$$\text{cross-stream roundtrip time } \cong \frac{2w}{c}\times\left(1+\frac{v^2}{2c^2}\right).$$

Looking at the two roundtrip times at the ends of the two paragraphs above, we see that they differ by an amount $(2w/c) \times v^2/2c^2$. Now, $2w/c$ is just the time the light would take if there were no aether wind at all, say, a few millionths of a second. If we take the aether windspeed to be equal to the earth's speed in orbit, for example, v/c is about 1/10,000, so v^2/c^2 is about 1/100,000,000. This means the time delay between the pulses reflected from the different mirrors reaching the telescope is about one-hundred-millionth of a few millionths of a second. It seems completely hopeless that such a short time delay could be detected. However, this turns out *not* to be the case, and Michelson was the first to figure out how to do it. The trick is to use the *interference* properties of the lightwaves. Instead of sending pulses of light, as we discussed above, Michelson sent in a steady beam of light of a single color. This can be visualized as a sequence of ingoing waves, with a wavelength one fifty-thousandth of an inch or so. Now this sequence of waves is split into two, and reflected as previously described. One set of waves goes upstream and downstream, the other goes across stream and back. Finally, they come together into the telescope and the eye. If the one that took longer is half a wavelength behind, its troughs will be on top of the crests of the first wave, they will cancel, and nothing will be seen. If the delay is less than that, there will still be some dimming. However, slight errors in the placement of the mirrors would have the same effect. This is one reason why the apparatus is built to be rotated. On turning it through 90 degrees, the upstream-downstream and the cross-stream waves change places. Now the other one should be behind. Thus, if there is an aether wind, if you watch through the telescope while you rotate the turntable, you should expect to see variations in the brightness of the incoming light.

To magnify the time difference between the two paths, in the actual experiment the light was reflected backwards and forwards several times, like a several lap race. For a diagram, click here. For an actual photograph of the real apparatus, click here.

Michelson calculated that an aether windspeed of only one or two miles a second would have observable effects in this experiment, so if the aether windspeed was comparable to the earth's speed in orbit around the sun, it would be easy to see. In fact, *nothing* was observed. The light intensity did not vary at all. Some time later, the experiment was redesigned so that an aether wind caused by the earth's daily rotation could be detected. Again, nothing was seen. Finally, Michelson wondered if the aether was somehow getting stuck to the earth, like the air in a below-decks cabin on a ship, so he redid the experiment on top of a high mountain in California. Again, no aether wind was observed. It was difficult to believe that the aether in the immediate vicinity of the earth was stuck to it and moving with it, because light rays from stars would deflect as they went from the moving faraway aether to the local stuck aether.

The only possible conclusion from this series of very difficult experiments was that the whole concept of an all-pervading aether was wrong from the start. Michelson was very reluctant to think along these lines. In fact, new theoretical insight into the nature of light had arisen in the 1860's from the brilliant theoretical work of Maxwell, who had written down a set of equations describing how electric and magnetic fields can give rise to each other. He had discovered that his equations predicted there could be waves made up of electric and magnetic fields, and the speed of these waves, deduced from experiments on how these fields link together, would be 186,300 miles per second. This is, of course, the speed of light, so it is natural to assume that light is made up of fast-varying electric and magnetic fields. But this leads to a big problem: Maxwell's equations predict a definite speed for light, and it *is* the speed found by measurements. But what is the speed to be measured relative to? The whole point of bringing in the aether was to give a picture for light resembling the one we understand for sound, compressional waves in a medium. The speed of sound through air is measured relative to air. If the wind is blowing towards you from the source of sound, you will hear the sound sooner. If there isn't an aether, though, this analogy doesn't hold up. So what does light travel at 186,300 miles per second relative to?

There is another obvious possibility, which is called the emitter theory: the light travels at 186,300 miles per second relative to the source of the light. The analogy here is between light emitted by a source and bullets emitted by a machine gun. The bullets come out at a definite speed (called the muzzle velocity) relative to the barrel of the gun. If the gun is mounted on the front of a tank, which is moving forward, and the gun is pointing forward, then relative to the ground the bullets are moving faster than they would if shot from a tank at rest. The simplest way to test the emitter theory of light, then, is to measure the speed of light emitted in the forward direction by a flashlight moving in the forward direction, and see if it exceeds the known speed of light by an amount equal to the speed of the flashlight. Actually, this kind of direct test of the emitter theory only became experimentally feasible in the nineteen-sixties. It is now possible to produce particles, called neutral pions, which decay each one in a little explosion, emitting a flash of light. It is also possible to have these pions moving forward at 185,000 miles per second when they self destruct, and to catch the light emitted in the forward direction, and clock its speed. It is found that, despite the expected boost from being emitted by a very fast source, the light from the little explosions is going forward at the usual speed of 186,300 miles per second. In the last century, the emitter theory was rejected because it was thought the appearance of certain astronomical phenomena, such as double stars, where two stars rotate around each other, would be affected. Those arguments have since been criticized, but the pion test is unambiguous. The definitive experiment was carried out by Alvager et al., *Physics Letters* **12**, 260 (1964).

Einstein's Answer

The results of the various experiments discussed above seem to leave us really stuck. Apparently light is not like sound, with a definite speed relative to some underlying medium. However, it is also not like bullets, with a definite speed relative to the source of the light. Yet when we measure its speed we always get the same result. How can all these facts be interpreted in a simple consistent way? We shall show how Einstein answered this question in the next lecture.

A detailed guide to setting up a Michelson-Morley experiment can be found at Nantes University.

Special Relativity

Galilean Relativity again

At this point in the course, we finally enter the twentieth century—Albert Einstein wrote his first paper on relativity in 1905. To put his work in context, let us first review just what is meant by "relativity" in physics. The first example, mentioned in a previous lecture, is what is called "Galilean relativity" and is nothing but Galileo's perception that by observing the motion of objects, alive or dead, in a closed room there is no way to tell if the room is at rest or is in fact in a boat moving at a steady speed in a fixed direction. (You *can* tell if the room is accelerating or turning around.) Everything looks the same in a room in steady motion as it does in a room at rest. After Newton formulated his Laws of Motion, describing how bodies move in response to forces and so on, physicists reformulated Galileo's observation in a slightly more technical, but equivalent, way: they said *the laws of physics are the same in a uniformly moving room as they are in a room at rest*. In other words, the same force produces the same acceleration, and an object experiencing no force moves at a steady speed in a straight line in either case. Of course, talking in these terms implies that we have clocks and rulers available so that we can actually time the motion of a body over a measured distance, so the physicist envisions the room in question to have calibrations along all the walls, so the position of anything can be measured, and a good clock to time motion. Such a suitably equipped room is called a "*frame of reference*"—the calibrations on the walls are seen as a frame which you can use to specify the precise position of an object at a given time. (This is the same as a set of "coordinates".) Anyway, the bottom line is that no amount of measuring of motions of objects in the "frame of reference" will tell you whether this is a frame at rest or one moving at a steady velocity.

What exactly do we mean by a frame "at rest" anyway? This seems obvious from our perspective as creatures who live on the surface of the earth—we mean, of course, at rest relative to fixed objects on the earth's surface. Actually, the earth's rotation means this isn't quite a fixed frame, and also the earth is moving in orbit at 18 miles per second. From an astronaut's point of view, then, a frame fixed relative to the sun might seem more reasonable. But why stop there? We believe the laws of physics are good throughout the universe. Let us consider somewhere in space far from the sun, even far from our galaxy. We would see galaxies in all directions, all moving in different ways. Suppose we now set up a frame of reference and check that Newton's laws still work. In particular, we check that the First Law holds—that a body experiencing no force moves at a steady speed in a straight line. *This First law is often referred to as The Principle of Inertia, and a frame in which it holds is called an Inertial Frame*. Then we set up another frame of reference, moving at a steady velocity relative to the first one, and find that Newton's laws are o.k. in this frame too. The point to notice here is that it is not at all obvious which—if either—of these frames is "at rest". We *can*, however, assert that they are both *inertial* frames, after we've checked that in both of them, a body with no forces acting on it moves at a steady speed in a straight line (the speed could be zero). In this situation, Michelson would have said that a frame "at rest" is one at rest relative to the aether. However, his own experiment found motion through the aether to be undetectable, so how would we ever know we were in the right frame?

As we mentioned in the last lecture, in the middle of the nineteenth century there was a substantial advance in the understanding of electric and magnetic fields. (In fact, this advance is in large part responsible for the improvement in living standards since that time.) The new understanding was summarized in a set of equations called Maxwell's equations describing how electric and magnetic fields interact and give rise to each other, just as, two centuries earlier, the new understanding of dynamics was summarized in the set of equations called Newton's laws. The important thing about Maxwell's equations for our present purposes is that they predicted waves made up of electric and magnetic fields that moved at 3×10^8 meters per second, and it was immediately realized that this was no coincidence—light waves must be nothing but waving electric and magnetic fields. (This is now fully established to be the case.)

It is worth emphasizing that Maxwell's work predicted the speed of light from the results of experiments that were not thought at the time they were done to have anything to do with light—experiments on, for example, the strength of electric field produced by waving a magnet. Maxwell was able to deduce a speed for waves like this using methods analogous to those by which earlier scientists had figured out the speed of sound from a knowledge of the density and the springiness of air.

Generalizing Galilean Relativity to Include Light: Special Relativity

We now come to Einstein's major insight: the Theory of Special Relativity. It is deceptively simple. Einstein first dusted off Galileo's discussion of experiments below decks on a uniformly moving ship, and restated it as :

The Laws of Physics are the same in all Inertial Frames.

Einstein then simply brought this up to date, by pointing out that the Laws of Physics must now include Maxwell's equations describing electric and magnetic fields as well as Newton's laws describing motion of masses under gravity and other forces. (*Note for experts and the curious*: we shall find that Maxwell's equations are completely unaltered by special relativity, but, as will become clear later, Newton's Laws do need a bit of readjustment to include special relativistic phenomena. The First Law is still o.k., the Second Law in the form $F = ma$ is not, because we shall find mass varies; we need to equate force to rate of change of momentum (Newton understood that, of course—that's the way he stated the law!). The Third Law, stated as action equals reaction, no longer holds because if a body moves, its electric field, say, does not readjust instantaneously—a ripple travels outwards at the speed of light. Before the ripple reaches another charged body, the electric forces between the two will be unbalanced. However, the crucial consequence of the Third Law—the conservation of momentum when two bodies interact, still holds. It turns out that the rippling field itself carries momentum, and everything balances.)

Demanding that Maxwell's equations be satisfied in all inertial frames has one major consequence as far as we are concerned. As we stated above, Maxwell's equations give the speed of light to be 3×10^8 meters per second. Therefore, *demanding that the laws of physics are the same in all inertial frames implies that the speed of any light wave, measured in any inertial frame, must be* 3×10^8 *meters per second*.

This then is the entire content of the Theory of Special Relativity: the Laws of Physics are the same in any inertial frame, and, in particular, any measurement of the speed of light in any inertial frame will always give 3×10^8 meters per second.

You *Really* Can't Tell You're Moving!

Just as Galileo had asserted that observing gnats, fish and dripping bottles, throwing things and generally jumping around would not help you to find out if you were in a room at rest or moving at a steady velocity, Einstein added that no kind of observation at all*, even measuring the speed of light across your room* to any accuracy you like, would help find out if your room was "really at rest". This implies, of course, that the concept of being "at rest" is meaningless. If Einstein is right, there is *no* natural rest-frame in the universe. Naturally, there can be no "aether", no thin transparent jelly filling space and vibrating with light waves, because if there were, *it* would provide the natural rest frame, and affect the speed of light as measured in other moving inertial frames as discussed above.

So we see the Michelson-Morley experiment was doomed from the start. There never was an aether wind. The light was not slowed down by going "upstream"—light *always* travels at the same speed, which we shall now call c,

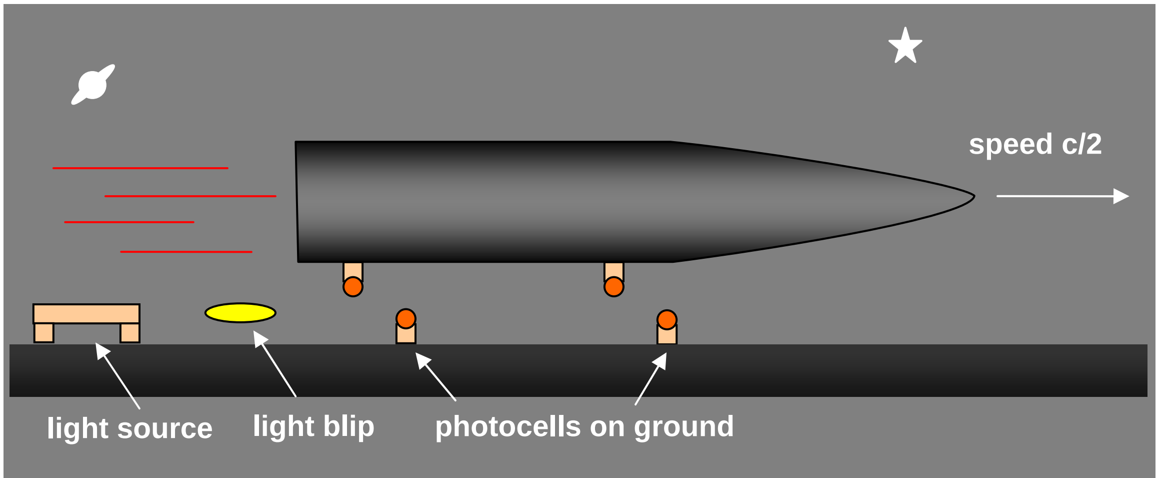
$$c = 3\times10^8 \text{ meters per second}$$

to save writing it out every time. *This now answers the question of what the speed of light, c, is relative to*. We already found that it is not like sound, relative to some underlying medium. It is also not like bullets, relative to the source of the light (the discredited emitter theory). *Light travels at c relative to the observer*, since if the observer sets up an inertial frame (clocks, rulers, etc.) to measure the speed of light he will find it to be c. (We always assume our observers are very competent experimentalists!)

Truth and Consequences

The Truth we are referring to here is the seemingly innocuous and plausible sounding statement that all inertial frames are as good as each other—the laws of physics are the same in all of them—and so the *speed of light* is the same in all of them. As we shall soon see, this Special Theory of Relativity has some surprising consequences, which reveal themselves most dramatically when things are moving at relative speeds comparable to the speed of light. Einstein liked to explain his theory using what he called “thought experiments” involving trains and other kinds of transportation moving at these speeds (technically unachievable so far!), and we shall follow his general approach.

To begin with, let us consider a simple measurement of the speed of light carried out at the same time in two inertial frames moving at half the speed of light relative to each other. The setup is as follows: on a flat piece of ground, we have a flashlight which emits a blip of light, like a strobe. We have two photocells, devices which click and send a message down a wire when light falls on them. The photocells are placed 10 meters apart in the path of the blip of light, they are somehow wired into a clock so that the time taken by the blip of light to travel from the first photocell to the second, in other words, the time between clicks, can be measured. From this time and the known distance between them, we can easily find the speed of the blip of light.



The speed of the same blip of light is measured by two observers, having relative speed c/2. Both measure the time the blip takes from one photocell to a second one 10 meters further on. Both find the speed to be c.

Meanwhile, there is another observer, passing overhead in a spaceship traveling at half the speed of light. She is also equipped with a couple of photocells, placed 10 meters apart on the bottom of her spaceship as shown, and she is able to measure the speed of the same blip of light, relative to her frame of reference (the spaceship). *The observer on the spaceship will measure the blip of light to be traveling at c relative to the spaceship, the observer on the ground will measure the same blip to be traveling at c relative to the ground*. That is the unavoidable consequence of the Theory of Relativity.

(*Note*: actually the picture above is not quite the way it would really look. As we shall find, objects moving at relativistic speeds are contracted, and this combined with the different times light takes to reach the eye from different parts of the ship would change the ship’s appearance. But this does not affect the validity of the statements above.)

Special Relativity: What Time is it?

Special Relativity in a Nutshell

Einstein's Theory of Special Relativity, discussed in the last lecture, may be summarized as follows:

The Laws of Physics are the same in any Inertial Frame of Reference. (Such frames move at steady velocities with respect to each other.)

These Laws include in particular Maxwell's Equations describing electric and magnetic fields, which predict that light always travels at a particular speed c, equal to about 3×10^8 meters per second, that is,186,300 miles per second.

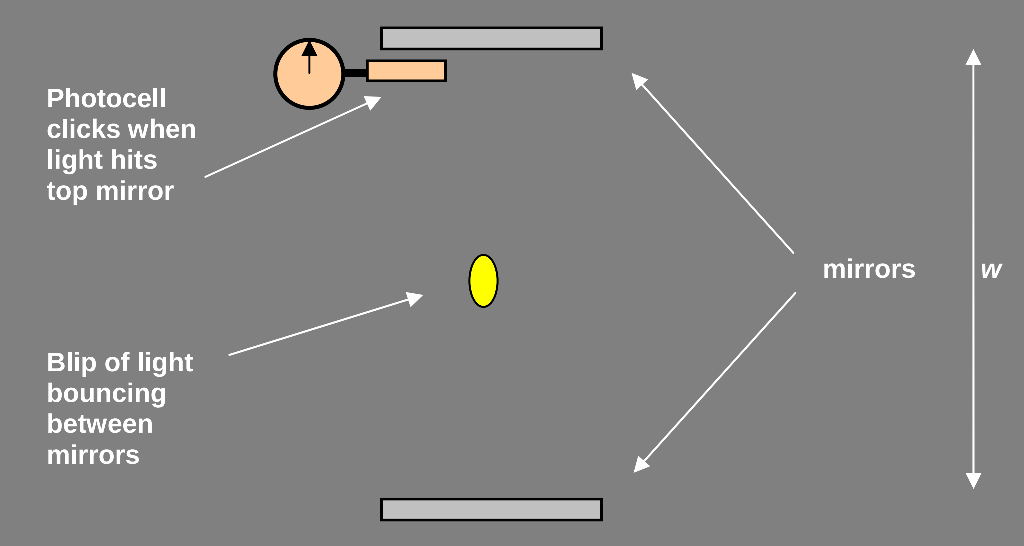
It follows that any measurement of the speed of any flash of light by any observer in any inertial frame will give the same answer c.

We have already noted one counter-intuitive consequence of this, that two different observers moving relative to each other, each measuring the speed of the *same* blob of light relative to himself, will *both* get c, even if their relative motion is in the same direction as the motion of the blob of light.

We shall now explore how this simple assumption changes everything we thought we understood about time and space.

A Simple but Reliable Clock

We mentioned earlier that each of our (inertial) frames of reference is calibrated (had marks at regular intervals along the walls) to measure distances, and has a clock to measure time. Let us now get more specific about the clock—we want one that is easy to understand in any frame of reference. Instead of a pendulum swinging back and forth, which wouldn't work away from the earth's surface anyway, we have a blip of light bouncing back and forth between two mirrors facing each other. We call this device a *light clock*. To really use it as a timing device we need some way to count the bounces, so we position a photocell at the upper mirror, so that it catches the edge of the blip of light. The photocell clicks when the light hits it, and this regular series of clicks drives the clock hand around, just as for an ordinary clock. Of course, driving the photocell will eventually use up the blip of light, so we also need some provision to reinforce the blip occasionally, such as a strobe light set to flash just as it passes and thus add to the intensity of the light. Admittedly, this may not be an easy way to build a clock, but the basic idea is simple.



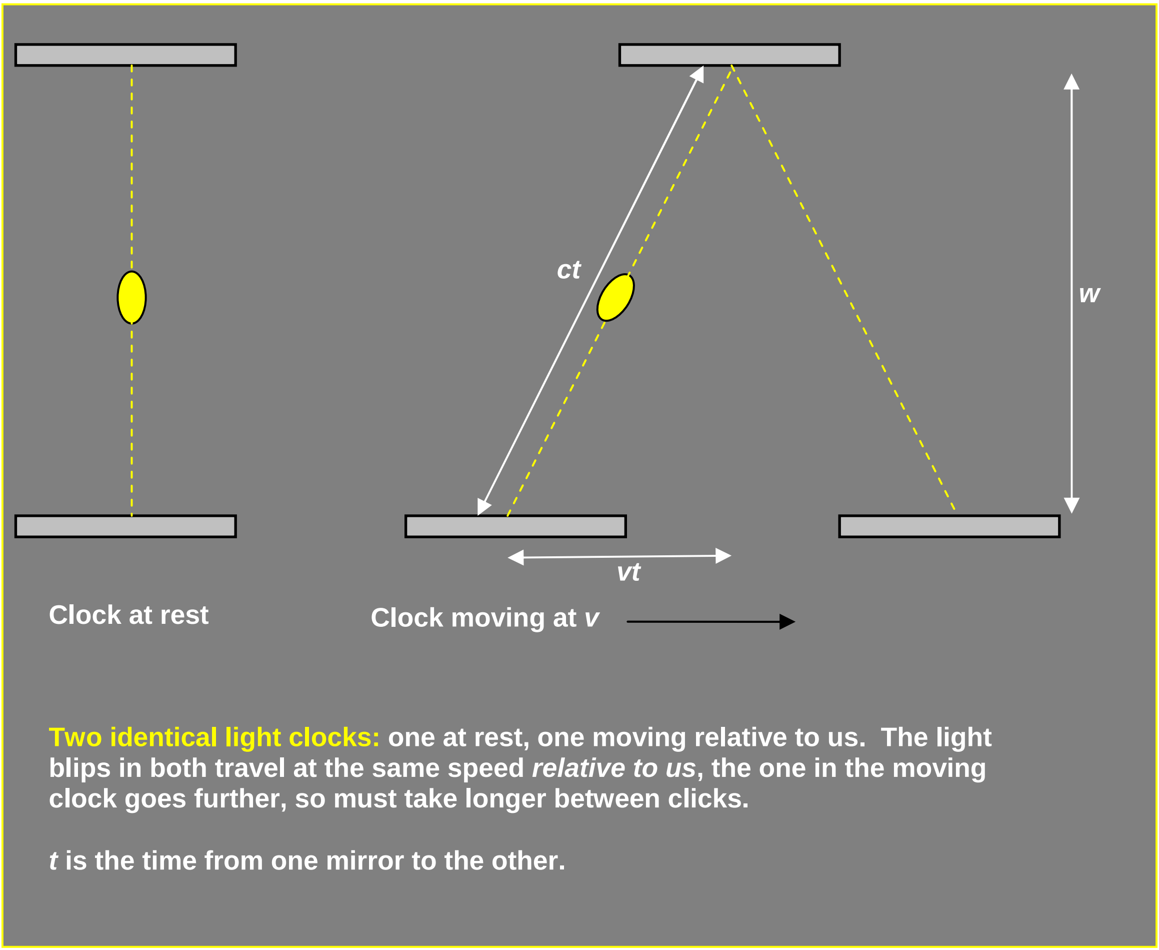
Einstein's Light Clock: a blip of light bounces between two parallel mirrors *w* meters apart; each time it hits the top mirror, the photocell moves the clock on one click.

Time between clicks: 2*w*/*c*.

It's easy to figure out how frequently our light clock clicks. If the two mirrors are a distance w apart, the round trip distance for the blip from the photocell mirror to the other mirror and back is $2w$. Since we know the blip always travels at c, we find the round trip time to be $2w/c$, so this is the time between clicks. This isn't a very long time for a reasonable sized clock! The crystal in a quartz watch "clicks " of the order of 10,000 times a second. That would correspond to mirrors about nine miles apart, so we need our clock to click about 1,000 times faster than that to get to a reasonable size. Anyway, let us assume that such purely technical problems have been solved.

Looking at Somebody Else's Clock

Let us now consider two observers, Jack and Jill, each equipped with a calibrated inertial frame of reference, and a light clock. To be specific, imagine Jack standing on the ground with his light clock next to a straight railroad line, while Jill and her clock are on a large flatbed railroad wagon which is moving down the track at a constant speed v. Jack now decides to check Jill's light clock against his own. He knows the time for his clock is $2w/c$ between clicks. Imagine it to be a slightly misty day, so with binoculars he can actually see the blip of light bouncing between the mirrors of Jill's clock. How long does he think that blip takes to make a round trip? The one thing he's sure of is that it must be moving at c = 186,300 miles per second, relative to him—that's what Einstein tells him. So to find the round trip time, all he needs is the round trip distance. This will *not* be $2w$, because the mirrors are on the flatbed wagon moving down the track, so, relative to Jack on the ground, when the blip gets back to the top mirror, that mirror has moved down the track some since the blip left, so the blip actually follows a zigzag path as seen from the ground.



Two identical light clocks: one at rest, one moving relative to us. The light blips in both travel at the same speed ***relative to us***, the one in the moving clock goes further, so must take longer between clicks.

t **is the time from one mirror to the other.**

Suppose now the blip in Jill's clock on the moving flatbed wagon takes time t to get from the bottom mirror to the top mirror as measured by Jack standing by the track. Then the length of the "zig" from the bottom mirror to the top mirror is necessarily ct, since that is the distance covered by any blip of light in time t. Meanwhile, the wagon has moved down the track a distance vt, where v is the speed of the wagon. This should begin to look familiar—it is precisely the same as the problem of the swimmer who swims at speed c relative to the water crossing a river flowing at v! We have again a right-angled triangle with hypotenuse ct, and shorter sides vt and w.

From Pythagoras, then,

$$c^2t^2 = v^2t^2 + w^2$$

so

$$t^2(c^2 - v^2) = w^2$$

or

$$t^2(1 - v^2/c^2) = w^2/c^2$$

and, taking the square root of each side, then doubling to get the round trip time, we conclude that Jack sees the time between clicks for Jill's clock to be:

$$\text{time between clicks for moving clock } = \frac{2w}{c}\frac{1}{\sqrt{1-v^2/c^2}}.$$

Of course, this gives the right answer $2w/c$ for a clock at rest, that is, $v = 0$.

This means that Jack sees Jill's light clock to be going slow—a longer time between clicks—compared to his own identical clock. Obviously, the effect is not dramatic at real railroad speeds. The correction factor is $\sqrt{1-v^2/c^2}$, which differs from 1 by about one part in a trillion even for a bullet train! Nevertheless, the effect is real and can be measured, as we shall discuss later.

It is important to realize that the only reason we chose a *light* clock, as opposed to some other kind of clock, is that its motion is very easy to analyze from a different frame. Jill could have a collection of clocks on the wagon, and would synchronize them all. For example, she could hang her wristwatch right next to the face of the light clock, and observe them together to be sure they always showed the same time. Remember, in her frame her light clock clicks every $2w/c$ seconds, as it is designed to do. Observing this scene from his position beside the track, Jack will see the synchronized light clock and wristwatch next to each other, and, of course, note that the wristwatch is *also* running slow by the factor $\sqrt{1-v^2/c^2}$. In fact, *all* her clocks, including her pulse, are slowed down by this factor according to Jack. Jill is aging more slowly because she's moving!

But this isn't the whole story—we must now turn everything around and look at it from Jill's point of view. *Her inertial frame of reference is just as good as Jack's.* She sees his light clock to be moving at speed v (backwards) so from her point of view *his* light blip takes the longer zigzag path, which means *his clock runs slower than hers.* That is to say, each of them will see the other to have slower clocks, and be aging more slowly. This phenomenon is called *time dilation*. It has been verified in recent years by flying very accurate clocks around the world on jetliners and finding they register less time, by the predicted amount, than identical clocks left on the ground. Time dilation is also very easy to observe in elementary particle physics, as we shall discuss in the next section.

Fitzgerald Contraction

Consider now the following puzzle: suppose Jill's clock is equipped with a device that stamps a notch on the track once a second. How far apart are the notches? From Jill's point of view, this is pretty easy to answer. She sees the track passing under the wagon at v meters per second, so the notches will of course be v meters apart. But Jack sees things differently. He sees Jill's clocks to be running slow, so he will see the notches to be stamped on the track at intervals of $1/\sqrt{1-v^2/c^2}$ seconds (so for a relativistic train going at $v = 0.8c$, the notches are stamped at intervals of 5/3 = 1.67 seconds). Since Jack agrees with Jill that the relative speed of the wagon and the track is v, he will assert the notches are not v meters apart, but $v/\sqrt{1-v^2/c^2}$ meters apart, a greater distance. Who is right? It turns out that Jack is right, because the notches are in his frame of reference, so he can wander over to them with a tape measure or whatever, and check the distance. This implies that as a result of her motion, Jill observes the notches to be closer together by a factor $\sqrt{1-v^2/c^2}$ than they would be at rest. This is called the *Fitzgerald contraction*, and applies not just to the notches, but also to the track and to Jack—everything looks somewhat squashed in the direction of motion!

Experimental Evidence for Time Dilation: Dying Muons

The first clear example of time dilation was provided over fifty years ago by an experiment detecting *muons*. These particles are produced at the outer edge of our atmosphere by incoming cosmic rays hitting the first traces of air. They are unstable particles, with a "half-life" of 1.5 microseconds (1.5 millionths of a second), which means that if at a given time you have 100 of them, 1.5 microseconds later you will have about 50, 1.5 microseconds after that 25, and so on. Anyway, they are constantly being produced many miles up, and there is a constant rain of them towards the surface of the earth, moving at very close to the speed of light. In 1941, a detector placed near the top of Mount Washington (at 6000 feet above sea level) measured about 570 muons per hour coming in. Now these muons are raining down from above, but dying as they fall, so if we move the detector to a lower altitude we expect it to detect fewer muons because a fraction of those that came down past the 6000 foot level will die before they get to a lower altitude detector. Approximating their speed by that of light, they are raining down at 186,300 miles per second, which turns out to be, conveniently, about 1,000 feet per microsecond. Thus they should reach the 4500 foot level 1.5 microseconds after passing the 6000 foot level, so, if half of them die off in 1.5 microseconds, as claimed above, we should only expect to register about 570/2 = 285 per hour with the same detector at this level. Dropping another 1500 feet, to the 3000 foot level, we expect about 280/2 = 140 per hour, at 1500 feet about 70 per hour, and at ground level about 35 per hour. (We have rounded off some figures a bit, but this is reasonably close to the expected value.)

To summarize: given the known rate at which these raining-down unstable muons decay, and given that 570 per hour hit a detector near the top of Mount Washington, we only expect about 35 per hour to survive down to sea level. In fact, when the detector was brought down to sea level, it detected about 400 per hour! How did they survive? The reason they didn't decay is that *in their frame of reference, much less time had passed*. Their actual speed is about 0.994c, corresponding to a time dilation factor of about 9, so in the 6 microsecond trip from the top of Mount Washington to sea level, their clocks register only 6/9 = 0.67 microseconds. In this period of time, only about one-quarter of them decay.

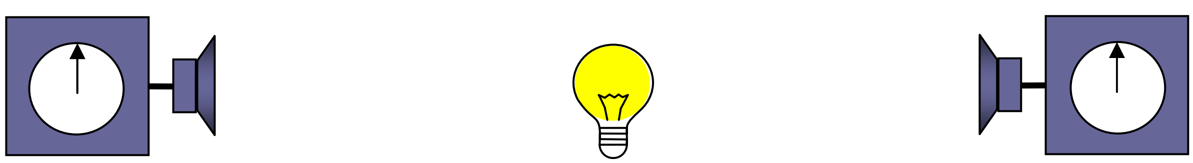
What does this look like from the muon's point of view? How do they manage to get so far in so little time? To them, Mount Washington and the earth's surface are approaching at 0.994c, or about 1,000 feet per microsecond. But in the 0.67 microseconds it takes them to get to sea level, it would seem that to them sea level could only get 670 feet closer, so how could they travel the whole 6000 feet from the top of Mount Washington? The answer is the Fitzgerald contraction. To them, Mount Washington is squashed in a vertical direction (the direction of motion) by a factor of $\sqrt{1-v^2/c^2}$, the same as the time dilation factor, which for the muons is about 9. So, to the muons, Mount Washington is only 670 feet high—this is why they can get down it so fast!

Special Relativity: Synchronizing Clocks

Suppose we want to synchronize two clocks that are some distance apart.

We could stand beside one of them and look at the other through a telescope, but we'd have to remember in that case that we are seeing the clock *as it was when the light left it*, and correct accordingly.

Another way to be sure the clocks are synchronized, assuming they are both accurate, is to start them together. How can we do that? We could, for example, attach a photocell to each clock, so when a flash of light reaches the clock, it begins running.



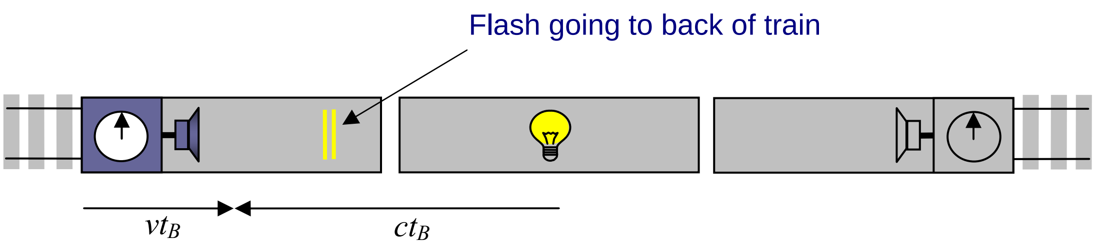
The clocks are triggered when the flash of light from the central bulb reaches the attached photocells.

If, then, we place a flashbulb at the midpoint of the line joining the two clocks, and flash it, the light flash will take the same time to reach the two clocks, so they will start at the same time, and therefore be synchronized.

Let us now put this whole arrangement - the two clocks and the midpoint flashbulb - on a train, and we suppose the train is moving at some speed v to the right, say half the speed of light or so.

Let's look carefully at the clock-synchronizing operation as seen from the ground. In fact, an observer on the ground would say the clocks are *not* synchronized by this operation! The basic reason is that he would see the flash of light from the middle of the train traveling at *c relative to the ground* in each direction, but he would also observe the back of the train coming at v to meet the flash, whereas the front is moving at v away from the bulb, so the light flash must go further to catch up.

In fact, it is not difficult to figure out how much later the flash reaches the front of the train compared with the back of the train, as viewed from the ground. First recall that as viewed from the ground the train has length $L\sqrt{1-v^2/c^2}$.



The train is moving to the right: the central bulb emits a flash of light. Seen from the ground, the part of the flash moving towards the rear travels at c, the rear travels at v to meet it.

Letting t_B be the time it takes the flash to reach the back of the train, it is clear from the figure that

$$vt_B + ct_B = \frac{L}{2}\sqrt{1-\frac{v^2}{c^2}}$$

from which t_B is given by

$$t_B = \frac{1}{c+v}\frac{L}{2}\sqrt{1-\frac{v^2}{c^2}}.$$

In a similar way, the time for the flash of light to reach the front of the train is (as measured by a ground observer)

$$t_F = \frac{1}{c-v}\frac{L}{2}\sqrt{1-\frac{v^2}{c^2}}.$$

Therefore the time difference between the starting of the two clocks, as seen from the ground, is

$$\begin{aligned}
t_F - t_B &= \left(\frac{1}{c-v} - \frac{1}{c+v}\right)\frac{L}{2}\sqrt{1-(v/c)^2} \\
&= \left(\frac{2v}{c^2-v^2}\right)\frac{L}{2}\sqrt{1-(v/c)^2} \\
&= \frac{2v}{c^2}\frac{1}{1-(v/c)^2}\frac{L}{2}\sqrt{1-(v/c)^2} \\
&= \frac{vL}{c^2}\frac{1}{\sqrt{1-(v/c)^2}}.
\end{aligned}$$

Remember, this is the time difference between the starting of the train's back clock and its front clock as measured by an observer on the ground with clocks on the ground. However, to this observer the clocks on the train appear to tick more slowly, by the factor $\sqrt{1-(v/c)^2}$, so that although the ground observer measures the time interval between the starting of the clock at the back of the train and the clock at the front as $\left(vL/c^2\right)\left(1/\sqrt{1-(v/c)^2}\right)$ seconds, he also sees the slow running clock at the back actually reading vL/c^2 seconds at the instant he sees the front clock to start.

To summarize: as seen from the ground, the two clocks on the train (which is moving at v in the x-direction) are running slowly, registering only $\sqrt{1-(v/c)^2}$ seconds for each second that passes. Equally important, the clocks—which are synchronized by an observer on the train—appear unsynchronized when viewed from the ground, the one at the back of the train reading vL/c^2 seconds ahead of the clock at the front of the train, where L is the rest length of the train (the length as measured by an observer on the train).

Note that if $L = 0$, that is, if the clocks are together, both the observers on the train and those on the ground will agree that they are synchronized. We need a *distance* between the clocks, as well as relative motion, to get a disagreement about synchronization.

The Lorentz Transformations

Problems with the Galilean Transformations

We have already seen that Newtonian mechanics is invariant under the *Galilean* transformations relating two inertial frames moving with relative speed v in the x-direction,

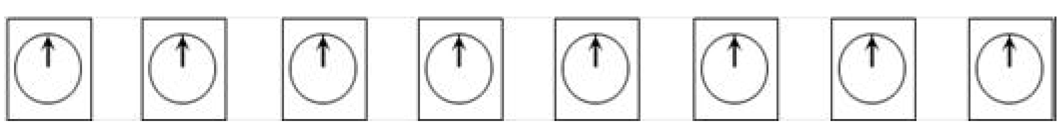
$$x = x' + vt'$$
$$y = y'$$
$$z = z'$$
$$t = t'.$$

However, these transformations presuppose that *time* is a well-defined universal concept, that is to say, it's the same time everywhere, and all observers can agree on what time it is. Once we accept the basic postulate of special relativity, however, that the laws of physics, including Maxwell's equations, are the same in all inertial frames of reference, and consequently the speed of light has the same value in all inertial frames, then as we have seen, observers in different frames do *not* agree on whether clocks some distance apart are synchronized. Furthermore, as we have discussed, measurements of moving objects are compressed in the direction of motion by the Lorentz-Fitzgerald contraction effect. Obviously, the above equations are too naïve! We must think more carefully about time and distance measurement, and construct new transformation equations consistent with special relativity.

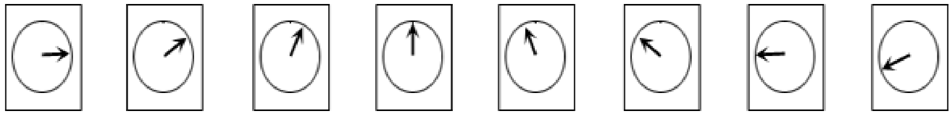
Our aim here, then, is to find a set of equations analogous to those above giving the coordinates of an *event* (x, y, z, t) in frame S, for example, a small bomb explosion, as functions of the coordinates (x ,y, z, t) of the same event measured in the parallel frame S' which is moving at speed v along the x-axis of frame S. Observers O at the origin in frame S and O' at the origin in frame S synchronize their clocks at $t = t' = 0$, at the instant they pass each other, that is, when the two frames coincide . (Using our previous notation, O is Jack and O' is Jill.)

To determine the time t' at which the bomb exploded in her frame, O' could determine the distance of the point (x', y', z') from the origin, and hence how long it would take light from the explosion to reach her at the origin. A more direct approach (which is helpful in considering transformations between different frames) is to imagine O' to have a multitude of helpers, with an array of clocks throughout the frame, which have all been synchronized by midpoint flashes as described in the previous lecture. Then the event—the bomb explosion—will be close to a clock, and that local clock determines the time t' of the event, so we do not need to worry about timing a light signal.

In frame S', then, O' and her crew have clocks all along the x'-axis (as well as everywhere else) and all synchronized:



Now consider how this string of clocks appears as viewed by O from frame S. First, since they are all moving at speed v, they will be registering time more slowly by the usual time dilation factor $\sqrt{1-v^2/c^2}$ than O's own physically identical clocks. Second, they will not be synchronized. From the clocks on a train argument in the last lecture, if the clocks are L apart as measured by O', successive clocks to the right (the direction of motion) will be behind by Lv/c^2 as observed by O.



It should be mentioned that this lack of synchronization as viewed from another frame only occurs for clocks *separated in the direction of relative motion*. Consider two clocks some distance apart **on the z' axis** of S'. If they are synchronized in S' by both being started by a flash of light from a bulb half way between them, it is clear that as viewed from S the light has to go the same distance to each of the clocks, so they will still be synchronized (although they will start later by the time dilation factor).

Deriving the Lorentz Transformations

Let us now suppose that O' and her crew observe a small bomb to explode in S' at $(x', 0, 0, t')$. In this section, we shall find the space coordinates and time (x, y, z, t) of this event as observed by O in the frame S. (As above, S' moves relative to S at speed v along the x-axis). In other words, we shall derive the Lorentz transformations—which are just the equations giving the four coordinates of an event in one inertial frame in terms of the coordinates of the same event in another inertial frame. We take y', z' zero because they transform trivially—there is no Lorentz contraction perpendicular to the motion, so $y = y'$ and $z = z'$.

First, we consider at what *time* the bomb explodes as measured by O. O''s crew found the bomb to explode at time t' as measured by a local clock, that is, one located at the site of the explosion, x'. Now, as observed by O from frame S, O''s clock at x' is *not* synchronized with O''s clock at the origin S'. When the bomb explodes and the clock at x' reads t', O will see O''s origin clock to read $t' + vx'/c^2$. What does O's own clock read at this point? Recall that O, O' synchronized their origin clocks at the moment they were together, at $t = t' = 0$. Subsequently, O will have observed O''s clock to be running slowly by the time-dilation factor. Therefore, when at the instant of the explosion he sees O''s origin clock to be reading $t' + vx'/c^2$, he will find that the true time t in his frame is equal to this appropriately scaled to allow for time dilation, that is,

$$t = \frac{t' + vx'/c^2}{\sqrt{1 - v^2/c^2}}.$$

This is the first of the Lorentz transformations.

The second question is: *where* does O observe the explosion to occur?

Since it occurs at time t after O' passed O, O' is vt meters beyond O at the time of the explosion. The explosion takes place x' meters beyond O', as measured by O', but of course O will see that distance x' as contracted to $x'\sqrt{1-v^2/c^2}$ since it's in a moving frame.

Therefore O observes the explosion at point x given by

$$x = vt + x'\sqrt{1-v^2/c^2}.$$

This can be written as an equation for x in terms of x', t' by substituting for t using the first Lorentz transformation above, to give

$$x = \frac{x'+vt'}{\sqrt{1-v^2/c^2}}.$$

Therefore, we have found the Lorentz transformations expressing the coordinates (x, y, z, t) of an event in frame S in terms of the coordinates (x', y', z', t') of the same event in frame S':

$$\begin{aligned}
x &= \frac{x'+vt'}{\sqrt{1-v^2/c^2}} \\
y &= y' \\
z &= z' \\
t &= \frac{t'+vx'/c^2}{\sqrt{1-v^2/c^2}}.
\end{aligned}$$

Notice that nothing in the above derivation depends on the x-velocity v of S' relative to S being positive. Therefore, the inverse transformation (from (x,y,z,t) to (x', y', z', t')) has exactly the same form as that given above with v replaced by $-v$.

Spheres of Light

Consider now the following scenario: suppose that as O' passes O (the instant both of them agree is at time $t' = t = 0$) O' flashes a bright light, which she observes to create an expanding spherical shell of light, centered on herself (imagine it' s a slightly foggy day, so she can see how the ripple of light travels outwards). At time t', then, O' (or, to be precise, her local observers out there in the frame) will see a shell of light of radius ct', that is to say, they will see the light to have reached all points (x', y', z') on the surface

$$x'^2 + y'^2 + z'^2 = c^2t'^2.$$

Question: how do O and his observers stationed throughout the frame S see this light as rippling outwards?

To answer this question, notice that the above equation for where the light is in frame S' at a particular time t' can be written

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = 0,$$

and can be thought of as *a surface in the four dimensional* (x', y', z', t') *space*, the totality of all the "events" of the light reaching any particular point. Now, to find the corresponding surface of events in the four dimensional (x, y, z, t) space, all we have to do is to change from one set of variables to the other using the Lorentz transformations:

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}} \\ y' &= y \\ z' &= z \\ t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}. \end{aligned}$$

On putting these values of (x', y', z', t') into $x'^2 + y'^2 + z'^2 - c^2t'^2 = 0,$ we find that the corresponding surface of events in (x, y, z, t) space is:

$$x^2 + y^2 + z^2 - c^2t^2 = 0.$$

This means that at time t, O and his observers in frame S will say the light has reached a spherical surface centered on O.

How can O' and O, as they move further apart, possibly both be right in maintaining that at any given instant the outward moving light pulse has a spherical shape, each saying it is centered on herself or himself?

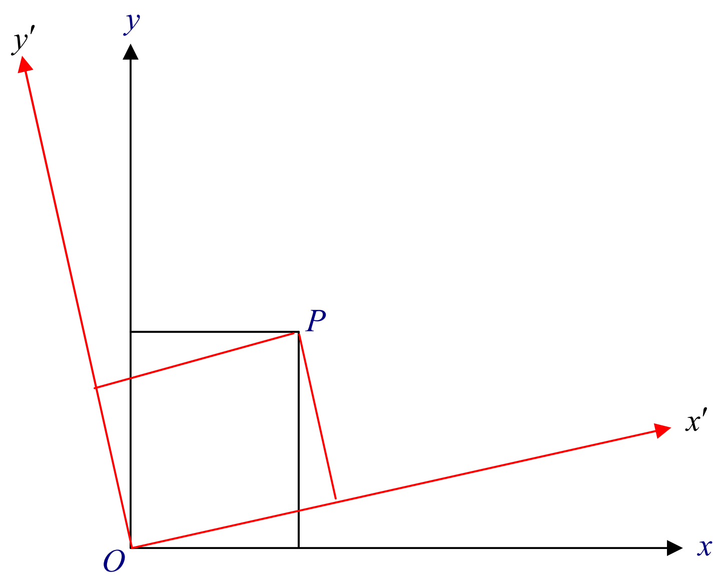
Imagine the light shell as O' sees it—at the instant t' she sees a sphere of radius r', in particular she sees the light to have reached the spots $+r'$ and $-r'$ on the x' axis. But from O''s point of view the expanding light sphere does *not* reach the point $+r'$ at the same time it reaches $-r'$! (This is just the old story of synchronizing the two clocks at the front and back of the train one more time.) That is why O does not see O''s sphere: *the arrival of the light at the sphere of radius r′ around O′ at time t′ corresponds in S to a continuum of different events happening at different times*.

Lorentz Invariants

We found above that for an event (x', y', z', t') for which $x'^2 + y'^2 + z'^2 - c^2t'^2 = 0,$ the coordinates of the event (x, y, z, t) as measured in the other frame S satisfy $x^2 + y^2 + z^2 - c^2t^2 = 0.$ The quantity $x^2 + y^2 + z^2 - c^2t^2$ is said to be a *Lorentz invariant*: it doesn't vary on going from one frame to another.

A simple two-dimensional analogy to this invariant is given by considering two sets of axes, Oxy and $Ox'y'$ having the same origin O, but the axis Ox' is at an angle to Ox, so one set of axes is the same as the other set but rotated. The point P with coordinates (x,y) has coordinates (x', y') measured on the $Ox'y'$ axes. The square of the distance of the point P from the common origin O is $x^2 + y^2$ and is also $x'^2 + y'^2$, so for the transformation from coordinates (x, y) to (x', y'), $x^2 + y^2$ is an invariant. Similarly, if a point P_1 has coordinates (x_1, y_1) and (x_1', y_1') and another point P_2 has coordinates (x_2, y_2) and (x_2', y_2') then clearly the two points are the same distance apart as measured with respect to the two sets of axes, so

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 = (x_1' - x_2')^2 + (y_1' - y_2')^2 .$$



This is really obvious: the distance between two points in an ordinary plane can't depend on the angle at which we choose to set our coordinate axes.

The Lorentz analog of this, dropping the y,z coordinates, can be written

$$c^2 (t_1 - t_2)^2 - (x_1 - x_2)^2 = c^2 (t_1' - t_2')^2 - (x_1' - x_2')^2 = s^2,$$

say, where s is some sort of measure of the "distance" between the two events (x_1, t_1) and (x_2, t_2).

This s is sometimes called the "space-time interval". *The big difference from the two-dimensional rotation case is that s^2 can be positive or negative.* For s^2 negative just taking s as its square root the "distance" itself would be imaginary—so we need to think this through carefully. As we shall see below, it's clear enough what to do in any particular case—the cases of spacelike and timelike separated events are best dealt with separately, at least to begin with.

Consider first two events simultaneous in frame S', so $t_1' = t_2'$. They will not be simultaneous in frame S, but they *will* satisfy

$$\left(x_1 - x_2\right)^2 - c^2\left(t_1 - t_2\right)^2 > 0.$$

We say the two events are *spacelike separated*. This means that they are sufficiently removed spatially that a light signal could not have time between them to get from one to the other, so one of these events could not be the *cause* of the other. The sequence of two events can be different in different frames if the events are spacelike separated. Consider again the starting of the two clocks at the front and back of a train as seen from the ground: the back clock starts first. Now imagine viewing this from a faster train overtaking the clock train—now the front clock will be the first to start. The important point is that although these events appear to occur in a different order in a different frame, neither of them could be the cause of the other, so cause and effect are not switched around.

Consider now two events which occur at the same place in frame S' at different times, $\left(x_1', t_1'\right)$ and $\left(x_2', t_1'\right)$. Then in frame S:

$$c^2\left(t_1 - t_2\right)^2 - \left(x_1 - x_2\right)^2 = c^2\left(t_1' - t_2'\right)^2 > 0.$$

These events are said to be *timelike* separated. There is no frame in which they are simultaneous. "Cause and effect" events are timelike separated.

The Light Cone

Let us try to visualize the surface in four-dimensional space described by the outgoing shell of light from a single flash,

$$x^2 + y^2 + z^2 - c^2t^2 = 0.$$

It is helpful to think about a simpler situation, the circular ripple spreading on the surface of calm water from a pebble falling in. Taking c here to be the speed of the water waves, it is easy to see that at time t after the splash the ripple is at

$$x^2 + y^2 - c^2t^2 = 0.$$

Now think about this as a surface in the three-dimensional space (x, y, t). The plane corresponding to time t cuts this surface in a circle of radius ct. This means the surface is a cone with its point at the origin. The four-dimensional space flash-of-light surface is not so easy to visualize, but is clearly the higher-dimensional analog: the plane surface corresponding to time t cuts it in a sphere instead of a circle. This surface is called the *lightcone*.

We have stated above that the separation of a point P (x, y, z, t) from the origin is spacelike if $x^2 + y^2 + z^2 - c^2t^2 > 0$ and timelike if $x^2 + y^2 + z^2 - c^2t^2 < 0$. It is said to be *lightlike* if $x^2 + y^2 + z^2 - c^2t^2 = 0$. Points on the light cone described above are lightlike separated from the origin. To be precise, the points corresponding to an outgoing shell of light from a flash at the origin at $t = 0$ form the *forward light cone*. Since the equation depends only on t^2, there is a solution with t negative, the "backward light cone", just the reflection of the forward light cone in the plane $t = 0$.

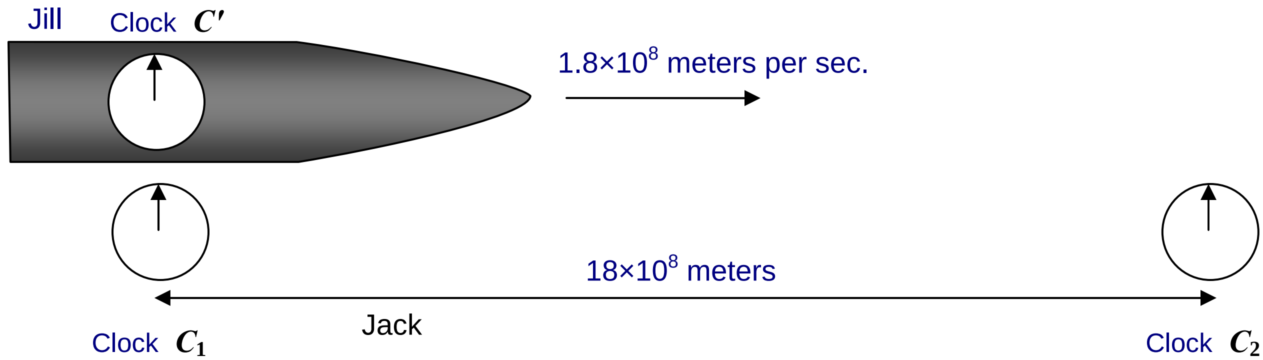
Possible causal connections are as follows: an event at the origin (0, 0, 0, 0) could cause an event in the forward light cone: so that is the "future", as seen from the origin. Events in the backward light cone—the "past"—could cause an event at the origin. There can be no causal link between an event at the origin and an event outside the light cones, since the separation is spacelike: outside the light cones is "elsewhere" as viewed from the origin.

Time Dilation: A Worked Example

"Moving Clocks Run Slow" *plus* "Moving Clocks Lose Synchronization" *plus* "Length Contraction" leads to consistency!

The object of this exercise is to show explicitly how it is possible for two observers in inertial frames moving relative to each other at a relativistic speed to each see the other's clocks as running slow and as being unsynchronized, and yet if they both look at the same clock at the same time from the same place (which may be far from the clock), they will *agree* on what time it shows!

Suppose that in Jack's frame we have two synchronized clocks C_1 and C_2 set 18 x 10^8 meters apart (that's about a million miles, or 6 light-seconds). Jill's spaceship, carrying a clock C', is traveling at $0.6c$, that is 1.8 x 10^8 meters per second, parallel to the line C_1C_2, passing close by each clock.



Jill in her relativistic rocket passes Jack's first clock at an instant when both their clocks read zero.

Suppose C' is synchronized with C_1 as they pass, so both read zero.

As measured by Jack the spaceship will take just 10 seconds to reach C_2, since the distance is 6 light seconds, and the ship is traveling at $0.6c$.

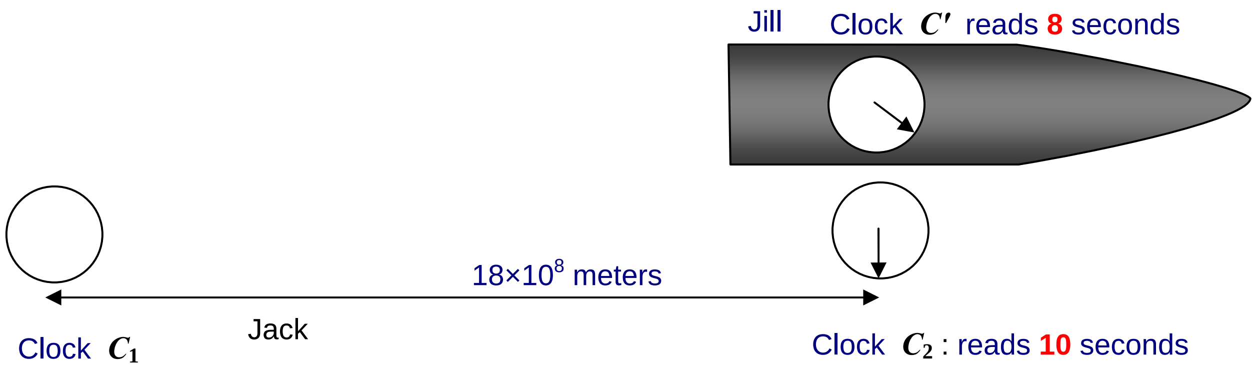
What does clock C' (the clock on the ship) read as it passes C_2?

The time dilation factor

$$\sqrt{1-\left(v^2/c^2\right)} = 4/5$$

so C', Jill's clock, will read 8 seconds.

Thus if both Jack and Jill are at C_2 as Jill and her clock C' pass C_2, both will agree that the clocks look like:



As Jill passes Jack's second clock, both see that his clock reads 10 seconds, hers reads 8 seconds.

How, then, can Jill claim that Jack's clocks C_1, C_2 are the ones that are running slow?

To Jill, C_1, C_2 *are* running slow, but remember they are *not synchronized*. To Jill, C_1 is *behind* C_2 by $Lv/c^2 = (L/c)\times(v/c) = 6\times 0.6 = 3.6$ seconds.

Therefore, Jill will conclude that since C_2 reads 10 seconds as she passes it, at that instant C_1 must be registering 6.4 seconds. Jill's own clock reads 8 seconds at that instant, *so she concludes that C_1 is running slow by the appropriate time dilation factor of 4/5*. This is how the change in synchronization makes it possible for both Jack and Jill to see the other's clocks as running slow.

Of course, Jill's assertion that as she passes Jack's second "ground" clock C_2 the first "ground" clock C_1 must be registering 6.4 seconds is not completely trivial to check! After all, that clock is now a million miles away!

Let us imagine, though, that both observers are equipped with Hubble-style telescopes attached to fast acting cameras, so reading a clock a million miles away is no trick.

To settle the argument, the two of them agree that as she passes the second clock, Jack will be stationed at the second clock, and at the instant of her passing they will both take telephoto digital snapshots of the faraway clock C_1, to see what time it reads.

Jack, of course, knows that C_1 is 6 light seconds away, and is synchronized with C_2 which at that instant is reading 10 seconds, so his snapshot must show C_1 to read 4 seconds. That is, looking at C_1 he sees it as it was six seconds ago.

What does Jill's digital snapshot show? It must be identical—two snapshots taken from the same place at the same time must show the same thing! So, Jill *must also* gets a picture of C_1 reading 4 seconds.

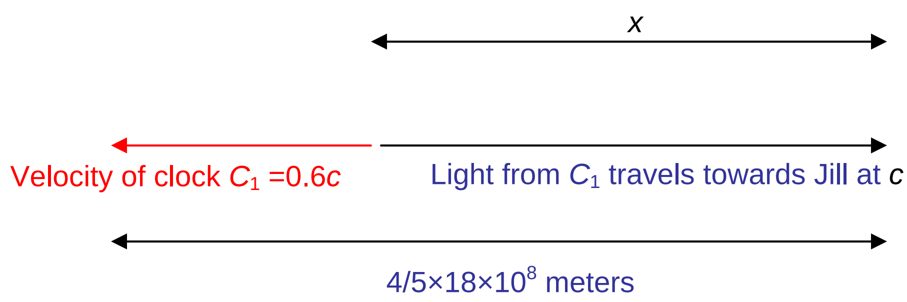
How can she reconcile a picture of the clock reading 4 seconds with her assertion that at the instant she took the photograph the clock was registering 6.4 seconds?

The answer is that she can if she knows her relativity!

First point: length contraction. To Jill, the clock C_1 is actually only 4/5 x 18 x 10^8 meters away (she sees the distance C_1C_2 to be Lorentz contracted!).

Second point: The light didn't even have to go that far! In her frame, the clock C_1 is *moving away*, so the light arriving when she's at C_2 must have left C_1 when it was closer—at distance x in the figure below. The figure shows the light in her frame moving from the clock towards her at speed c, while at the same time the clock itself is moving to the left at 0.6c.

It might be helpful to imagine yourself in her frame of reference, so you are at rest, and to think of clocks C_1 and C_2 as being at the front end and back end respectively of a train that is going past you at speed 0.6c. Then, at the moment the back of the train passes you, you take a picture (through your telescope, of course) of the clock at the front of the train. Obviously, the light from the front clock that enters your camera at that instant left the front clock some time ago. During the time that light traveled towards you at speed c, the front of the train itself was going in the opposite direction at speed 0.6c. But you know the length of the train in your frame is 4/5 x 18 x 10^8 meters, so since at the instant you take the picture the back of the train is passing you, the front of the train must be 4/5 x 18 x 10^8 meters away. Now that distance, 4/5 x 18 x 10^8, is the sum of the distance the light entering your camera traveled plus the distance the train traveled in the same time, that is, (1 + 0.6)/1 times the distance the light traveled.



As Jill passes C_2, she photographs C_1: at that instant, she knows C_1 is $4/5 \times 18 \times 10^8$ meters away in her frame, but the light reaching her camera at that moment left C_1 when it was at a distance x, not so far away. As the light traveled towards her at speed c, C_1 was moving away at a speed of 0.6c, so the distance $4/5 \times 18 \times 10^8$ meters is the *sum* of how far the light traveled towards her and how far the clock traveled away from her, both starting at x.

So the image of the first ground clock she sees and records as she passes the second ground clock must have been emitted when the first clock was a distance x from her in her frame, where

$$x(1+3/5) = 4/5 \times 18 \times 10^8 \text{ meters, so } x = 9 \times 10^8 \text{ meters.}$$

Having established that the clock image she is seeing as she takes the photograph left the clock when it was only 9 x 10^8 meters away, that is, 3 light seconds, she concludes that she is observing the first ground clock as it was three seconds ago.

Third point: time dilation. The story so far: she has a photograph of the first ground clock that shows it to be reading 4 seconds. She knows that the light took three seconds to reach her. So, what can she conclude the clock must actually be registering at the instant the photo was taken? If you are tempted to say 7 seconds, you have forgotten that in her frame, the clock is moving at 0.6c and hence *runs slow* by a factor 4/5.

Including the time dilation factor correctly, she concludes that in the 3 seconds that the light from the clock took to reach her, the clock itself will have ticked away 3 × 4/5 seconds, or 2.4 seconds.

Therefore, since the photograph shows the clock to read 4 seconds, and she finds the clock must have run a further 2.4 seconds, she deduces that at the instant she took the photograph the clock must actually have been registering 6.4 seconds, which is what she had claimed all along!

The key point of this lecture is that at first it seems impossible for two observers moving relative to each other to both maintain that the other one's clocks run slow. However, by bringing in the other necessary consequences of the theory of relativity, the Lorentz contraction of lengths, and that clocks synchronized in one frame are out of synchronization in another by a precise amount that follows necessarily from the constancy of the speed of light, the whole picture becomes completely consistent!

More Relativity: The Train and the Twins

Einstein's Definition of Common Sense

As you can see from the lectures so far, although Einstein's theory of special relativity solves the problem posed by the Michelson-Morley experiment—the nonexistence of an ether—it is at a price. The simple assertion that the speed of a flash of light is always c in any inertial frame leads to consequences that defy common sense. When this was pointed out somewhat forcefully to Einstein, his response was that common sense is the layer of prejudices put down before the age of eighteen. All our intuition about space, time and motion is based on childhood observation of a world in which no objects move at speeds comparable to that of light. Perhaps if we had been raised in a civilization zipping around the universe in spaceships moving at relativistic speeds, Einstein's assertions about space and time would just seem to be common sense. The real question, from a scientific point of view, is not whether special relativity defies common sense, but whether it can be shown to lead to a *contradiction*. If that is so, common sense wins. Ever since the theory was published, people have been writing papers claiming it *does* lead to contradictions. The previous lecture, the worked example on time dilation, shows how careful analysis of an apparent contradiction leads to the conclusion that in fact there was no contradiction after all. In this lecture, we shall consider other apparent contradictions and think about how to resolve them. This is the best way to build up an understanding of relativity.

Trapping a Train in a Tunnel

One of the first paradoxes to be aired was based on the Fitzgerald contraction. Recall that any object moving relative to an observer will be seen by that observer to be contracted, foreshortened in the direction of motion by the ubiquitous factor $\sqrt{1-v^2/c^2}$. Einstein lived in Switzerland, a very mountainous country where the railroads between towns often go through tunnels deep in the mountains.

Suppose a train of length L is moving along a straight track at a relativistic speed and enters a tunnel, also of length L. There are bandits inhabiting the mountain above the tunnel. They observe a short train, one of length $L\sqrt{1-v^2/c^2}$, so they wait until this short train is completely inside the tunnel of length L, then they close doors at the two ends, and the train is trapped fully inside the mountain. Now look at this same scenario from the point of view of someone on the train. He sees a train of length L, approaching a tunnel of length $L\sqrt{1-v^2/c^2}$, so the tunnel is not as long as the train from his viewpoint! What does he think happens when the bandits close both the doors?

The Tunnel Doors are Closed Simultaneously

The key to understanding what is happening here is that we said the bandits closed the two doors at the ends of the tunnel *at the same time*. How could they arrange to do that, since the doors are far apart? They could use walkie-talkies, which transmit radio waves, or just flash a light down the tunnel, since it's long and straight. Remember, though, that the train is itself going at a speed close to that of light, so they have to be quite precise about this timing! The simplest way to imaging them synchronizing the closings of the two doors is to assume they know the train's timetable, and at a prearranged appropriate time, a light is flashed halfway down the tunnel, and the end doors are closed when the flash of light reaches the ends of the tunnel. Assuming the light was positioned correctly in the middle of the tunnel, that should ensure that the two doors close simultaneously.

Or are they?

Now consider this door-closing operation from the point of view of someone on the train. Assume he's in an observation car and has incredible eyesight, and there's a little mist, so he actually sees the light flash, and the two flashes traveling down the tunnels towards the two end doors. Of course*, the train is a perfectly good inertial frame*, so he sees these two flashes to be traveling in opposite directions, but *both at c, relative to the train*. Meanwhile, he sees the tunnel itself to be moving rapidly relative to the train. Let us say the train enters the mountain through the "front" door. The observer will see the door at the other end of the tunnel, the "back" door, to be rushing towards him, and rushing to meet the flash of light. Meanwhile, once he's in the tunnel, the front door is receding rapidly behind him, so the flash of light making its way to that door has to travel further to catch it. So the two flashes of light going down the tunnel in opposite directions do not reach the two doors simultaneously as seen from the train.

The concept of simultaneity, events happening at the same time, is not *invariant as we move from one inertial frame to another*. The man on the train sees the back door close first, and, if it is not quickly reopened, the front of the train will pile into it before the front door is closed behind the train.

Does the Fitzgerald Contraction Work Sideways?

The above discussion is based on Einstein's prediction that objects moving at relativistic speed appear shrunken in their direction of motion. How do we know that they're not shrunken in all three directions, i.e. moving objects maybe keep the same shape, but just get smaller? This can be seen *not* to be the case through a symmetry argument, also due to Einstein. Suppose two trains traveling at equal and opposite relativistic speeds, one north, one south, pass on parallel tracks. Suppose two passengers of equal height, one on each train, are standing leaning slightly out of open windows so that their noses should very lightly touch as they pass each other. Now, if *N* (the northbound passenger) sees *S* as shrunken in height, *N*'s nose will brush against *S*'s forehead, say, and *N* will feel *S*'s nose brush his chin. Afterwards, then, *N* will have a bruised chin (plus nose), *S* a bruised forehead (plus nose). But this is a perfectly symmetric problem, so *S* would say *N* had the bruised forehead, etc. They can both get off their trains at the next stations and get together to check out bruises. They must certainly be symmetrical! The only *consistent symmetrical solution* is given by asserting that *neither* sees the other to shrink in height (i.e. in the direction perpendicular to their relative motion), so that their noses touch each other. Therefore, the Lorentz contraction *only* operates in the direction of motion, objects get squashed but not shrunken.

How to Give Twins Very Different Birthdays

Perhaps the most famous of the paradoxes of special relativity, which was still being hotly debated in national journals in the fifties, is the twin paradox. The scenario is as follows. One of two twins - the sister—is an astronaut . (Flouting tradition, we will take fraternal rather than identical twins, so that we can use "he" and "she" to make clear which twin we mean). She sets off in a relativistic spaceship to alpha-centauri, four light-years away, at a speed of, say, 0.6c. When she gets there, she immediately turns around and comes back. As seen by her brother on earth, her clocks ran slowly by the time dilation factor $\sqrt{1-v^2/c^2}$, so although the round trip took 8/0.6 years = 160 months by earth time, she has only aged by 4/5 of that, or 128 months. So as she steps down out of the spaceship, she is 32 months younger than her twin brother.

But wait a minute—how does this look from *her* point of view? She sees the earth to be moving at 0.6c, first away from her then towards her. So she must see her brother's clock on earth to be running slow! So doesn't she expect her brother on earth to be the younger one after this trip?

The key to this paradox is that this situation is not as symmetrical as it looks. The two twins have quite different experiences. The one on the spaceship is *not* in an inertial frame during the initial acceleration *and* the turnaround and braking periods. (To get an idea of the speeds involved, to get to 0.6c at the acceleration of a falling stone would take over six months.) Our analysis of how a clock in one inertial frame looks as viewed from another doesn't work during times when one of the frames isn't inertial - in other words, when one is accelerating.

The Twins Stay in Touch

To try to see just how the difference in ages might develop, let us imagine that the twins stay in touch with each other throughout the trip. Each twin flashes a powerful light once a month, according to their calendars and clocks, so that by counting the flashes, each one can monitor how fast the other one is aging.

The questions we must resolve are:

If the brother, on earth, flashes a light once a month, how frequently, as measured by her clock, does the sister see his light to be flashing as she moves away from earth at speed 0.6c?

How frequently does she see the flashes as she is *returning* at 0.6*c*?

How frequently does the brother on earth see the flashes from the spaceship?

Once we have answered these questions, it will be a matter of simple bookkeeping to find how much each twin has aged.

Figuring the Observed Time between Flashes

To figure out how frequently each twin observes the other's flashes to be, we will use some results from the previous lecture, on time dilation. In some ways, that was a very small scale version of the present problem. Recall that we had two "ground" clocks only one million miles apart. As the astronaut, conveniently moving at 0.6*c*, passed the first ground clock, both that clock and her own clock read zero. As she passed the second ground clock, her own clock read 8 seconds and the *first* ground clock, which she photographed at that instant, she observed to read 4 seconds.

That is to say, after 8 seconds had elapsed on her own clock, constant *observation* of the first ground clock would have revealed it to have registered only 4 seconds. (This effect is compounded of time dilation and the fact that as she moves away, the light from the clock is taking longer and longer to reach her.)

Our twin problem is the same thing, at the same speed, but over a longer time - we conclude that observation of any earth clock from the receding spacecraft will reveal it to be running *at half speed*, so the brother's flashes will be seen at the spacecraft to arrive every two months, by spacecraft time.

Symmetrically, as long as the brother on earth observes his sister's spacecraft to be moving away at 0.6*c*, he will see light from her flashes to be arriving at the earth every two months by earth time.

To figure the frequency of her brother's flashes observed as she returns towards earth, we have to go back to our previous example and find how the astronaut traveling at 0.6*c* observes time to be registered by the *second* ground clock, the one she's approaching.

We know that as she passes that clock, it reads 10 seconds and her own clock reads 8 seconds. We must figure out what she would have seen that second ground clock to read had she glanced at it through a telescope as she passed the first ground clock, at which point both her own clock and the first ground clock read zero. But at that instant, the reading she would see on the second ground clock must be the same as would be seen by an observer on the ground, standing by the first ground clock and observing the second ground clock through a telescope. Since the ground observer knows both ground clocks are synchronized, and the first ground clock reads zero, and the second is 6 light seconds distant, it must read -6 seconds if observed at that instant.

Hence the astronaut will observe the second ground clock to progress from -6 seconds to +10 seconds during the period that her own clock goes from 0 to 8 seconds. In other words, she sees the clock she is approaching at $0.6c$ to be running *at double speed*.

Finally, back to the twins. During her journey back to earth, the sister will see the brother's light flashing twice a month. (Evidently, the time dilation effect does not fully compensate for the fact that each succeeding flash has less far to go to reach her.)

We are now ready to do the bookkeeping: first, from the sister's point of view.

What does she see?

At $0.6c$, she sees the distance to alpha-centauri to be contracted by the familiar $\sqrt{1-v^2/c^2}=0.8$ to a distance of 3.2 light years, which at $0.6c$ will take her a time 5.333 years, or, more conveniently, 64 months. During the outward trip, then, she will see 32 flashes from home, she will see her brother to age by 32 months.

Her return trip will also take 64 months, during which time she will see 128 flashes, so over the whole trip she will see 128 + 32 = 160 flashes. That means she will have seen her brother to age by 160 months or 13 years 4 months.

What does he see?

As he watches for flashes through his telescope, the stay-at-home brother will see his sister to be aging at half his own rate of aging as long as he sees her to be moving away from him, then aging at twice his rate as he sees her coming back. At first glance, this sounds the same as what she sees—but it isn't! The important question to ask is *when* does he see her turn around? To him, her outward journey of 4 light years' distance at a speed of $0.6c$ takes her 4/0.6 years, or 80 months. BUT he doesn't *see* her turn around until 4 years later, because of the time light takes to get back to earth from alpha-centauri! In other words, he will actually see her aging at half his rate for 80 + 48 = 128 months, during which time he will see 64 flashes.

When he *sees* his sister turn around, she is already more than half way back! Remember, in his frame the whole trip takes 160 months (8 light years at $0.6c$) so he will only see her aging at twice his rate during the last 160 - 128 = 32 months, during which period he will see all 64 flashes she sent out on her return trip.

Therefore, by counting the flashes of light she transmitted once a month, he will conclude she has aged 128 months on the trip, which by his clock and calendar took 160 months. So when she steps off the spacecraft 32 months younger than her twin brother, neither of them will be surprised!

The Doppler Effect

The above analysis hinges on the fact that a traveler approaching a flashing light at $0.6c$ will see it flashing at *double* its "natural" rate - the rate observed by someone standing still with the light - and a traveler receding at $0.6c$ from a flashing light will see it to flash at only *half* its natural rate.

This is a particular example of the *Doppler Effect*, first discussed in 1842 by the German physicist Christian Doppler. There is a Doppler Effect for sound waves too. Sound is generated by a vibrating object sending a succession of pressure pulses through the air. These pressure waves are analogous to the flashes of light. If you are approaching a sound source you will encounter the pressure waves more frequently than if you stand still. This means you will hear a higher frequency sound. If the distance between you and the source of sound is increasing, you will hear a lower frequency. This is why the note of a jet plane or a siren goes lower as it passes you. The details of the Doppler Effect for sound are a little different than those for light, because the speed of sound is not the same for all observers - it's 330 meters per second relative to the air.

It isn't difficult to find the general formula for the Doppler shift, that is, the change in frequency observed when the source of waves (or periodic signals in general) is moving. For example, consider a light flashing once a second as observed by someone in the same frame as the light. Let us imagine the light to be attached to a spaceship, passing us at a relativistic speed v, and imagine we see a flash at the instant it passes us. When will we see the next flash? First, the spaceship's clock is running slow according to us, so it will take $1/\sqrt{1-v^2/c^2}$ seconds before it emits the next flash. But it's also moving away from us, so we won't see that next flash until the light has traveled back to us over the distance covered by the spaceship between flashes. Since the spaceship is traveling at v, and the time between flashes as measured in our frame is $1/\sqrt{1-v^2/c^2}$, the distance the spaceship covers between flashes measures in our frame is $v/\sqrt{1-v^2/c^2}$. Since the light coming back to us from the flash is traveling at c, it covers this distance in time $v/c\sqrt{1-v^2/c^2}$.

Thus the total time between our observing the first flash as the spaceship passes close by us and the second flash emitted one second later by the spaceship clock is:

$$\frac{1}{\sqrt{1-v^2/c^2}}+\frac{v}{c\sqrt{1-v^2/c^2}}=\sqrt{\frac{1+v/c}{1-v/c}}$$

The change in frequency is the inverse of the change in time between flashes. Also, notice that the shift for a spaceship *approaching* at speed v is the *inverse* of that for one *receding* at speed v. We already established that above for the special case of $v = 0.6c$, where we found the frequency halved for the spaceship receding, doubled for it approaching.

There are some significant differences between the Doppler effect for light and that for sound. The most obvious one is that if something is approaching you at a speed higher than the speed of sound, you won't hear a thing until it hits! As we shall see, nothing can approach you at greater than the speed of light. Another difference is that the Doppler shift for light depends only on the velocity of the emitter relative to that of the receiver. For sound, the shift depends on the velocities of both of them relative to the air. Finally, consider a signal from a moving object moving along a straight line, but not towards you. For example, consider a train moving along a straight track, and the nearest point on the track is one mile from where you're standing. What frequency do you hear (or see) for the signal emitted when the train was at the nearest point to you on the track? For a sound signal, you would hear no Doppler shifting at this point. For a light signal, however, you *would* see a frequency shift downwards equal to the time dilation factor. This is called the *transverse* Doppler shift, and is famous historically because it was first detected in 1938 by two experimenters, Ives and Stillwell, who even then didn't believe in special relativity! They interpreted their result as the slowing down of a clock as it moved through the aether.

An important astronomical application of the Doppler Effect is the *red shift*. The light from very distant galaxies is redder than the light from similar galaxies nearer to us. This is because the further away a galaxy is, the faster it is moving away from us, as the Universe expands. The light is redder because red light is low frequency light (blue is high) and we see low frequency light for the same reason that the astronaut receding from earth sees flashes less frequently. In fact, the farthest away galaxies we can see are receding faster than the 0.6*c* of our astronaut!

Adding Velocities: A Walk on the Train

The Formula

If I walk from the back to the front of a train at 3 m.p.h., and the train is traveling at 60 m.p.h., then common sense tells me that my speed relative to the ground is 63 m.p.h. As we have seen, this obvious truth, the simple addition of velocities, follows from the Galilean transformations. Unfortunately, it can't be quite right for high speeds! We know that for a flash of light going from the back of the train to the front, the speed of the light relative to the ground is exactly the same as its speed relative to the train, not 60 m.p.h. different. Hence it is necessary to do a careful analysis of a fairly speedy person moving from the back of the train to the front as viewed from the ground, to see how velocities *really* add.

We consider our standard train of length *L* moving down the track at steady speed *v*, and equipped with synchronized clocks at the back and the front. The walker sets off from the back of the train when that clock reads zero. Assuming a steady walking speed of *u* meters per second (relative to the train, of course), the walker will see the front clock to read *L/u* seconds on arrival there.

How does this look from the ground? Let's assume that at the instant the walker began to walk from the clock at the back of the train, the back of the train was passing the ground observer's clock, and both these clocks (one on the train and one on the ground) read zero. The ground observer sees the walker reach the clock at the front of the train at the instant that clock reads L/u (this is in agreement with what is observed *on* the train—two simultaneous events *at the same place* are simultaneous to all observers), but at this same instant, the ground observer says the train's *back* clock, where the walker began, reads $L/u + Lv/c^2$. (This follows from our previously established result that two clocks synchronized in one frame, in which they are L apart, will be out of synchronization in a frame in which they are moving at v along the line joining them by a time Lv/c^2.)

Now, how much time elapses as measured by the ground observer's clock during the walk? At the instant the walk began, the ground observer saw the clock at the back of the train (which was right next to him) to read zero. At the instant the walk ended, the ground observer would say that clock read $L/u + Lv/c^2$, from the paragraph above. But the ground observer would see that clock to be running slow, by the usual time dilation factor: so he would measure the time of the walk on his own clock to be:

$$\frac{L/u+Lv/c^2}{\sqrt{1-\left(v^2/c^2\right)}}.$$

How *far* does the walker move as viewed from the ground? In the time t_W, the train travels a distance vt_W, so the walker moves this distance plus the length of the train. Remember that the train is contracted as viewed from the ground! It follows that the distance covered *relative to the ground* during the walk is:

$$\begin{aligned}
d_W &= vt_W + L\sqrt{1-\left(v^2/c^2\right)} \\
&= v\frac{L/u+Lv/c^2}{\sqrt{1-\left(v^2/c^2\right)}} + L\sqrt{1-\left(v^2/c^2\right)} \\
&= \frac{vL/u+Lv^2/c^2+L-L\left(v^2/c^2\right)}{\sqrt{1-\left(v^2/c^2\right)}} \\
&= \frac{L\left(1+v/u\right)}{\sqrt{1-\left(v^2/c^2\right)}}.
\end{aligned}$$

The walker's *speed* relative to the ground is simply d_W/t_W, easily found from the above expressions:

$$\frac{d_W}{t_W} = \frac{1+v/u}{1/u+v/c^2} = \frac{u+v}{1+uv/c^2}.$$

This is the appropriate formula for adding velocities. Note that it gives the correct answer, $u + v$, in the low velocity limit, and also if u or v equals c, the sum of the velocities is c.

Exercise 1: The direct derivation given above is really equivalent to a rederivation of the Lorentz equations. The equation describing the walker's path on the train is just $x' = ut'$. Substitute this in the Lorentz equations and prove it leads to a path relative to the ground given by $x = wt$, with w given by the velocity addition formula.

Exercise 2: Suppose a spaceship is equipped with a series of one-shot rockets, each of which can accelerate the ship to $c/2$ from rest. It uses one rocket to leave the solar system (ignore gravity here) and is then traveling at $c/2$ (relative to us) in deep space. It now fires its second rocket, keeping the same direction. Find how fast it is moving relative to us. It now fires the third rocket, keeping the same direction. Find its new speed. Can you draw any general conclusions from your results?

Walking *Across* the Train

Imagine now a rather wide train, of width w, and the walker begins the walk across the train, which is now equipped with clocks on both sides, when the clock where he begins reads $t = 0$. For walking speed $u_{y'}$ (relative to the train, and across the train is the y-direction) when he reaches the clock at the other side it will read $w/u_{y'}$. How is this seen from the ground? The width of the train w will be the same, there is no Lorentz contraction in the y-direction for motion in the x-direction. The beginning and ending clocks will also be synchronized as seen from the ground, since they are separated in the y-direction but not the x-direction. However, they are clocks moving at relativistic speed, so they will exhibit the familiar time dilation factor. That is, when they read $w/u_{y'}$, a clock on the ground will read

$$\frac{w}{u_{y'}}\frac{1}{\sqrt{1-v^2/c^2}}.$$

Thus, as observed from the ground, walking directly across the train is slowed down by the time dilation factor, just as is every other activity on the train as seen from the ground.

However, for steady motion on the train in an arbitrary direction, velocity components $(u_{x'}, u_{y'})$ the cross-train velocity transforms in a more complicated way, because the train clocks at the beginning and end of the walk are now separated in the x-direction, so if they register an elapsed time of w/u_y a ground observer would add a lack of synchronicity term

$$\frac{Lv}{c^2} = \frac{wu_{x'}}{u_{y'}}\frac{v}{c^2}.$$

Thus the time for the walk as observed from the ground

$$t_W = \left(\frac{w}{u_{y'}} + \frac{wu_{x'}}{u_{y'}} \frac{v}{c^2} \right) / \sqrt{1 - v^2 / c^2} = \frac{w}{u_y}.$$

From this we find the general formula for transformation of transverse velocities:

$$u_y = u_{y'} \frac{\sqrt{1 - v^2 / c^2}}{1 + u_{x'} v / c^2}.$$

For the special case of walking directly across the train, $u_{x'} = 0$, we recover the earlier result, that transverse velocity is simply slowed by the time dilation effect.

Testing the Addition of Velocities Formula

Actually, the first test of the addition of velocities formula was carried out in the 1850s! Two French physicists, Fizeau and Foucault, measured the speed of light in water, and found it to be c/n, where n is the refractive index of water, about 1.33. (This was the result predicted by the wave theory of light.)

They then measured the speed of light (relative to the ground) in *moving* water, by sending light down a long pipe with water flowing through it at speed v. They discovered that the speed relative to the ground was not just $v + c/n$, but had an extra term, $v + c/n - v/n^2$. Their (incorrect) explanation was that the light was a complicated combination of waves in the water and waves in the aether, and the moving water was only partially dragging the aether along with it, so the light didn't get the full speed v of the water added to its original speed c/n.

The true explanation of the extra term is much simpler: velocities don't simply add. To add the velocity v to the velocity c/n, we must use the addition of velocities formula above, which gives the light velocity relative to the ground to be:

$$(v + c/n)/(1 + v/nc)$$

Now, v is much smaller than c or c/n, so $1/(1 + v/nc)$ can be written as $(1 - v/nc)$, giving:

$$(v + c/n)(1 - v/nc)$$

Multiplying this out gives $v + c/n - v/n^2 - v/n \times v/c$, and the last term is smaller than v by a factor v/c, so is clearly negligible.

Therefore, the 1850 experiment looking for "aether drag" in fact confirms the relativistic addition of velocities formula! Of course, there are many other confirmations. For example, any velocity added to c still gives c. Also, it indicates that the speed of light is a speed limit for all objects, a topic we shall examine more carefully in the next lecture.

Relativistic Dynamics

The Story So Far: A Brief Review

The first coherent statement of what physicists now call *relativity* was Galileo's observation almost four hundred years ago that if you were in a large closed room, you could not tell by observing how things move-living things, thrown things, dripping liquids-whether the room was at rest in a building, say, or below decks in a large ship moving with a steady velocity. More technically (but really saying the same thing!) we would put it that *the laws of motion are the same in any inertial frame.* That is, these laws really only describe *relative* positions and velocities. In particular, they do not single out a special inertial frame as the one that's "really at rest". This was later all written down more formally, in terms of *Galilean transformations*. Using these simple linear equations, motion analyzed in terms of positions and velocities in one inertial frame could be translated into any other. When, after Galileo, Newton wrote down his Three Laws of Motion, they were of course invariant under the Galilean transformations, and valid in any inertial frame.

About two hundred years ago, it became clear that light was not just a stream of particles (as Newton had thought) but manifested definite *wavelike* properties. This led naturally to the question of what, exactly, was waving, and the consensus was that space was filled with an *aether*, and light waves were ripples in this all-pervading aether analogous to sound waves in air. Maxwell's discovery that the equations describing electromagnetic phenomena had wavelike solutions, and predicted a speed which coincided with the measured speed of light, suggested that electric and magnetic fields were stresses or strains in the aether, and Maxwell's equations were presumably only precisely correct in the frame in which the aether was at rest. However, very precise experiments which should have been able to detect this aether all failed.

About a hundred years ago, Einstein suggested that maybe *all* the laws of physics were the same in *all* inertial frames, generalizing Galileo's pronouncements concerning motion to include the more recently discovered laws of electricity and magnetism. This would imply there could be no special "really at rest" frame, even for light propagation, and hence no aether. This is a very appealing and very simple concept: the same laws apply in all frames. What could be more reasonable? As we have seen, though, it turns out to clash with some beliefs about space and time deeply held by everybody encountering this for the first time. The central prediction is that since the speed of light follows from the laws of physics (Maxwell's equations) and some simple electrostatic and magnetostatic experiments, which are clearly frame-independent, *the speed of light is the same in all inertial frames*. That is to say, the speed of a *particular flash* of light will always be measured to be $3\ 10^8$ meters per second even if measured by different observers moving rapidly relative to each other, where each observer measures the speed of the flash relative to himself. Nevertheless, experiments have show again and again that Einstein's elegant insight is right, and everybody's deeply held beliefs are wrong.

We have discussed in detail the *kinematical* consequences of Einstein's postulate: how measurements of position, time and velocity in one frame relate to those in another, and how apparent paradoxes can be resolved by careful analysis. So far, though, we have not thought much about dynamics. We know that Newton's Laws of Motion were invariant under the Galilean transformations between inertial frames. We now know that the Galilean transformations are in fact *incorrect* except in the low speed nonrelativistic limit. Therefore, we had better look carefully at Newton's Laws of Motion in light of our new knowledge.

Newton's Laws Revisited

Newton's First Law, the Principle of Inertia, that an object subject to no external forces will continue to move in a straight line at steady speed, is equally valid in special relativity. Indeed, it is the defining property of an inertial frame that this *is* true, and the content of special relativity is transformations between such frames.

Newton's Second Law, stated in the form force = mass x acceleration, cannot be true as it stands in special relativity. This is evident from the formula we derived for addition of velocities. Think of a rocket having many stages, each sufficient to boost the remainder of the rocket (including the unused stages) to $c/2$ from rest. We could fire them one after the other in a carefully timed way to generate a continuous large force on the rocket, which would get it to $c/2$ in the first firing. If the acceleration continued, the rocket would very soon be exceeding the speed of light. Yet we know from the addition of velocities formula that in fact the rocket never reaches c. Evidently, Newton's Second Law needs updating.

Newton's Third Law, action = reaction, also has problems. Consider some attractive force between two rapidly moving bodies. As their distance apart varies, so does the force of attraction. We might be tempted to say that the force of A on B is the opposite of the force of B on A, at each instant of time, but that implies *simultaneous* measurements at two bodies some distance from each other, and if it happens to be true in A's inertial frame, it won't be in B's.

Conservation Laws

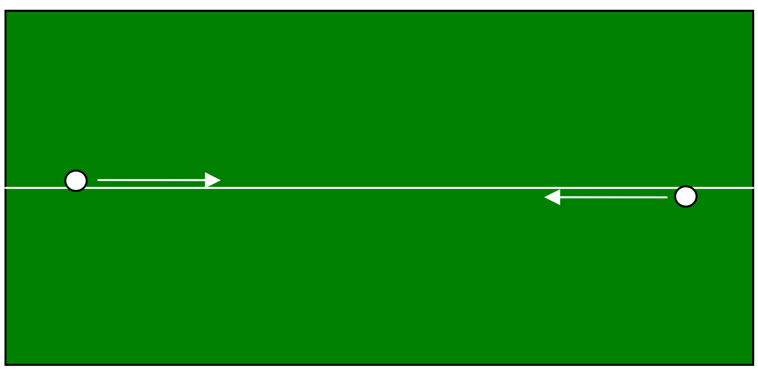
In nonrelativistic Newtonian physics, the Third Law tells us that two interacting bodies feel equal but opposite forces from the interaction. Therefore from the Second Law, the rate of change of momentum of one of the bodies is equal and opposite to that of the other body, thus the *total* rate of change of momentum of the system caused by the interaction is zero. Consequently, for any *closed* dynamical system (no outside forces acting) *the total momentum never changes*. This is the *law of conservation of momentum*. It does *not* depend on the details of the forces of interaction between the bodies, only that they be equal and opposite.

The other major dynamical conservation law is the conservation of energy. This was not fully formulated until long after Newton, when it became clear that frictional heat generation, for example, could quantitatively account for the apparent loss of kinetic plus potential energy in actual dynamical systems.

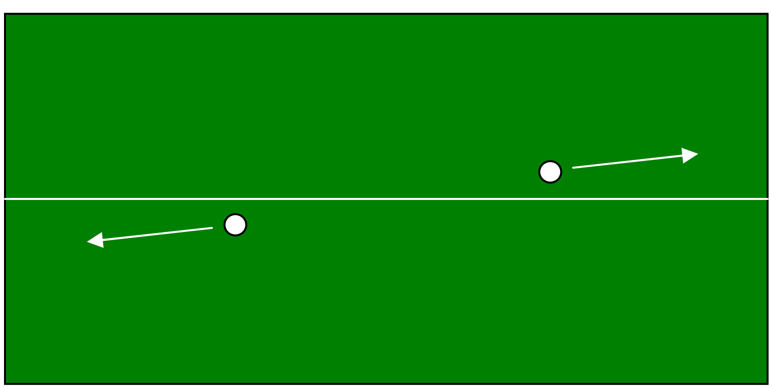
Although these conservation laws were originally formulated within a Newtonian worldview, their very general nature suggested to Einstein that they might have a wider validity. Therefore, as a working hypothesis, he assumed them to be satisfied *in all inertial frames*, and explored the consequences. We follow that approach.

Momentum Conservation on the Pool Table

As a warm-up exercise, let us consider conservation of momentum for a collision of two balls on a pool table. We draw a chalk line down the middle of the pool table, and shoot the balls close to, but on opposite sides of, the chalk line from either end, at the same speed, so they will hit in the middle with a glancing blow, which will turn their velocities through a small angle. In other words, if initially we say their (equal magnitude, opposite direction) velocities were parallel to the *x*-direction—the chalk line—then after the collision they will also have equal and opposite small velocities in the *y*-direction. (The *x*-direction velocities will have decreased very slightly).



Balls on pool table moving towards glancing collision



Motion of balls on table after collision

A Symmetrical Spaceship Collision

Now let us repeat the exercise on a grand scale. Suppose somewhere in space, far from any gravitational fields, we set out a string one million miles long. (It could be between our two clocks in the time dilation experiment). This string corresponds to the chalk line on the pool table. Suppose now we have two identical spaceships approaching each other with equal and opposite velocities parallel to the string from the two ends of the string, aimed so that they suffer a slight glancing collision when they meet in the middle. It is evident from the symmetry of the situation that momentum is conserved in both directions. In particular, the rate at which one spaceship moves away from the string after the collision - its y-velocity - is equal and opposite to the rate at which the other one moves away from the string.

But now consider this collision as observed by someone in one of the spaceships, call it A. (Remember, momentum must be conserved in *all* inertial frames—they are all equivalent—there is nothing special about the frame in which the string is at rest.) Before the collision, he sees the string moving very fast by the window, say a few meters away. After the collision, he sees the string to be moving away, at, say, 15 meters per second. This is because spaceship A has picked up a velocity perpendicular to the string of 15 meters per second. Meanwhile, since this is a completely symmetrical situation, an observer on spaceship B would certainly deduce that her spaceship was moving away from the string at 15 meters per second as well.

Just how symmetrical is it?

The crucial question is: *how fast does an observer in spaceship A see spaceship B to be moving away from the string?* Let us suppose that relative to spaceship A, spaceship B is moving away (in the x-direction) at $0.6c$. First, recall that distances perpendicular to the direction of motion are not Lorentz contracted. Therefore, when the observer in spaceship B says she has moved 15 meters further away from the string in a one second interval, the observer watching this movement from spaceship A will agree on the 15 meters - but disagree on the one second! He will say her clocks run slow, so as measured by his clocks 1.25 seconds will have elapsed as she moves 15 meters in the y-direction.

It follows that, as a result of time dilation, this collision as viewed from spaceship A does *not* cause equal and opposite velocities for the two spaceships in the y-direction. Initially, both spaceships were moving parallel to the x-axis - there was zero momentum in the y-direction. Consider y-direction momentum conservation in the inertial frame in which A was initially at rest. An observer in that frame measuring y-velocities after the collision will find A to be moving at 15 meters per second, B to be moving at -0.8 x 15 meters per second in the y-direction. So how can we argue there is zero total momentum in the y-direction *after* the collision, when the identical spaceships do *not* have equal and opposite velocities?

Einstein rescues Momentum Conservation

Einstein was so sure that momentum conservation must always hold that he rescued it with a bold hypothesis: the mass of an object must depend on its speed! In fact, the mass must increase with speed in just such a way as to cancel out the lower *y*-direction velocity resulting from time dilation. That is to say, if an object at rest has a mass m_0, moving at a speed v it must have mass

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

to conserve *y*-direction momentum.

Note that this is an undetectably small effect at ordinary speeds, but as an object approaches the speed of light, the mass increases without limit!

Of course, we have taken a very special case here: a particular kind of collision. The reader might well wonder if the same mass correction would work in other types of collision, for example a straight line collision in which a heavy object rear-ends a lighter object. The algebra is straightforward, if tedious, and it is found that this mass correction factor does indeed ensure momentum conservation for any collision in all inertial frames.

Mass Really *Does* Increase with Speed

Deciding that masses of objects must depend on speed like this seems a heavy price to pay to rescue conservation of momentum! However, it is a prediction that is not difficult to check by experiment. The first confirmation came in 1908, measuring the mass of fast electrons in a vacuum tube. In fact, the electrons in an old-fashioned color TV tube are about half a percent heavier than electrons at rest, and this must be allowed for in calculating the magnetic fields used to guide them to the screen.

Much more dramatically, in modern particle accelerators very powerful electric fields are used to accelerate electrons, protons and other particles. It is found in practice that these particles become heavier and heavier as the speed of light is approached, and hence need greater and greater forces for further acceleration. Consequently, the speed of light is a natural absolute speed limit. Particles are accelerated to speeds where their mass is thousands of times greater than their mass measured at rest, usually called the "rest mass".

Warning: It should be mentioned that some people don't like the statement that mass increases with speed, they feel that the word "mass" should be restricted to the rest mass of an object, which we've called m_0. This difference of definition has no physical content, however—it's just a matter of taste. We would write momentum as $p = mv$, they would write our m_0 as m, and say the formula for momentum in their notation is $p = mv/\sqrt{1 - v^2/c^2}$. Either way, a fast electron is that much harder to deflect from a straight line.

Mass and Energy Conservation: Kinetic Energy and Mass for Very Fast Particles

As everyone has heard, in special relativity mass and energy are not separately conserved, in certain situations mass m can be converted to energy $E = mc^2$. This equivalence is closely related to the mass increase with speed, as we shall see. Suppose a constant force F accelerates a particle of rest mass m_0 in a straight line. The work done by the force in accelerating the particle as it travels a distance d is Fd, and this work has given the particle kinetic energy.

As a warm up, recall the elementary derivation of the kinetic energy $\frac{1}{2}mv^2$ of an ordinary *non-relativistic* (i.e. slow moving) object of mass m. Suppose it starts from rest. Then after time t, it has traveled distance $d = \frac{1}{2}at^2$, and $v = at$. From Newton's second law, $F = ma$, the work done by the force $Fd = mad = \frac{1}{2}ma^2t^2 = \frac{1}{2}mv^2$.

This won't work if the mass is varying, because Newton's Second Law isn't always $F = ma$, for variable mass it's

$$F = dp/dt,$$

force = rate of change of momentum, and if the mass changes the momentum changes, even at constant velocity.

An instructive extreme case is the kinetic energy of a particle traveling close to the speed of light, as particles do in accelerators. In this regime, the change of speed with increasing momentum is negligible! Instead,

$$F = \frac{dp}{dt} = \frac{d(mv)}{dt} \cong \frac{dm}{dt}c$$

where as usual c is the speed of light. This is what happens in a particle accelerator for a charged particle in a constant electric field, with $F = qE$.

Since the particle is moving at a speed very close to c, in time dt it will move cdt and the force will do work $Fcdt$. The equation above can be rewritten

$$Fcdt = (dm)c^2$$

So the energy dE expended by the accelerating force in the time dt yields an increase in mass, and $dE = (dm)c^2$. Provided the speed is close to c, this can of course be integrated to an excellent approximation, to relate a finite particle mass change to the energy expended in accelerating it.

Kinetic Energy and Mass for Slow Particles

Recall that to get momentum to be conserved in all inertial frames, we had to assume an increase of mass with speed by the factor $1/\sqrt{1-v^2/c^2}$. This necessarily implies that *even a slow-moving object has a tiny mass increase if it is put in motion.*

How does this mass increase relate to the kinetic energy? Consider a mass m_0, moving at speed v, much less than the speed of light. Its kinetic energy $E = \frac{1}{2}m_0v^2$, as discussed above. Its mass is $m_0/\sqrt{1-v^2/c^2}$, which we can write as $m_0 + dm$, so dm is the tiny mass increase we know must occur. It's easy to calculate dm.

For $v/c \ll 1$, we can make the approximations

$$\sqrt{1-v^2/c^2} \cong 1 - \tfrac{1}{2}v^2/c^2$$

and

$$\frac{1}{1-\frac{1}{2}v^2/c^2} \cong 1 + \tfrac{1}{2}v^2/c^2.$$

So, for $v/c \ll 1$,

$$\begin{aligned} m(v) &\cong m_0\left(1+\tfrac{1}{2}v^2/c^2\right) \\ dm &\cong \left(\tfrac{1}{2}m_0v^2\right)/c^2 = KE/c^2. \end{aligned}$$

Again, the mass increase dm is related to the kinetic energy KE by $KE = (dm)c^2$. Having looked at two simple cases, we're ready to derive the general result, valid over the whole range of possible speeds.

Kinetic Energy and Mass for Particles of Arbitrary Speed

We have shown in the two sections above that (in the two limiting cases) when a force does work to increase the kinetic energy of a particle it also causes the mass of the particle to increase by an amount equal to the increase in energy divided by c^2. In fact this result is exactly true over the whole range of speed from zero to arbitrarily close to the speed of light, as we shall now demonstrate. For a particle of rest mass m_0 accelerating along a straight line (from rest) under a constant force F,

$$
\begin{aligned}
F &= \frac{d}{dt}(mv) \\
&= \frac{dm}{dt}v + m\frac{dv}{dt} \\
&= \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}}\frac{v^2}{c^2}\frac{dv}{dt} + \frac{m_0}{\left(1-v^2/c^2\right)^{1/2}}\frac{dv}{dt} \\
&= \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}}\frac{dv}{dt}.
\end{aligned}
$$

Therefore, the work done when the particle moves a distance dx is

$$
\begin{aligned}
Fdx &= \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}}\frac{dv}{dt}dx \\
&= \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}}vdv,
\end{aligned}
$$

using $v = dx/dt$.

Therefore the total work done from rest—the kinetic energy—is:

$$
\int Fdx = \int \frac{m_0}{\left(1-v^2/c^2\right)^{3/2}}vdv = \left(m - m_0\right)c^2.
$$

(The integral is easily done by making the substitution $y = v^2/c^2$.)

So we see that in the general case the work done on the body, by definition its kinetic energy, is just equal to its mass increase multiplied by c^2.

To understand why this isn't noticed in everyday life, try an example, such as a jet airplane weighing 100 tons moving at 2,000mph. 100 tons is 100,000 kilograms, 2,000mph is about 1,000 meters per second. That's a kinetic energy ½ mv^2 of ½.10^{11}joules, but the corresponding mass change of the airplane down by the factor $c^2 = 9.10^{16}$, giving an actual mass increase of about half a milligram, not too easy to detect!

Notation: *m* and m_0

As stated earlier, we use m_0 to denote the "rest mass" of an object, and m to denote its relativistic mass, $m = m_0/\sqrt{1-v^2/c^2}$.

In this notation, we follow French and Feynman. *Krane and Tipler, in contrast, use m for the rest mass*. Using m as we do gives neater formulas for momentum and energy, but is not without its dangers. One must remember that m is *not* a constant, but a function of speed. Also, *one must remember* that the relativistic kinetic energy is $(m-m_0)c^2$, and *not* equal to $\frac{1}{2}mv^2$, even with the relativistic mass!

Example: take $v^2/c^2 = 0.99$, find the kinetic energy, and compare it with $\frac{1}{2}mv^2$ (using the relativistic mass).

Mass and Energy

Rest Energy

The fact that feeding energy into a body increases its mass suggests that the mass m_0 of a body *at rest*, multiplied by c^2, can be considered as a quantity of energy. The truth of this is best seen in interactions between elementary particles. For example, there is a particle called a *positron* which is exactly like an electron except that it has positive charge. If a positron and an electron collide at low speed (so there is very little kinetic energy) they both disappear in a flash of electromagnetic radiation. This can be detected and its energy measured. It turns out to be $2m_0c^2$ where m_0 is the mass of the electron (and the positron).

Thus particles can "vaporize" into pure energy, that is, electromagnetic radiation. The energy m_0c^2 of a particle at rest is called its "rest energy". Note, however, that an electron can only be vaporized by meeting with a positron, and there are very few positrons around normally, for obvious reasons-they just don't get far. (Although occasionally it has been suggested that some galaxies may be antimatter!)

Einstein's Box

An amusing "experiment" on the equivalence of mass and energy is the following: consider a closed box with a flashlight at one end and light-absorbing material at the other end. Imagine the box to be far out in space away from gravitational fields or any disturbances. Suppose the light flashes once, the flash travels down the box and is absorbed at the other end.

Einstein's Box: far out in space, a flash of light emitted by a bulb attached at one end is absorbed at the other end.

Now it is known from Maxwell's theory of electromagnetic waves that a flash of light carrying energy E also carries momentum $p = E/c$. Thus, as the flash leaves the bulb and goes down the tube, the box recoils, like a gun, to conserve overall momentum. Suppose the whole apparatus has mass M and recoils at velocity v. Of course, $v << c$.

Then from conservation of momentum in the frame in which the box was initially at rest:

$$Mv = E / c,$$

the recoil momentum of the box equals (minus) the momentum of the flash emitted.

After a time $t = L/c$ the light hits the far end of the tube, is absorbed, and the whole thing comes to rest again. (We are assuming that the distance moved by the box is tiny compared to its length.)

How far did the box move? It moved at speed v for time t, so it moved distance

$$d = vt = vL / c.$$

From the conservation of momentum equation above, we see that $v = E/Mc$, so the distance d the box moved over is:

$$d = \frac{EL}{mc^2}.$$

Now, the important thing is that there are *no* external forces acting on this system, so *the center of mass cannot have moved*!

The only way this makes sense is to say that to counterbalance the mass M moving d backwards, the light energy *must have transferred a small mass m*, say, the length L of the tube so that

$$Md = mL$$

and balance is maintained. From our formula for d above, we can figure out the necessary value of m,

$$m = \frac{M}{L} d = \frac{M}{L} \frac{EL}{Mc^2} = \frac{E}{c^2}$$

so

$$E = mc^2.$$

We have therefore established that *transfer of energy implies transfer of the equivalent mass*. Our only assumptions here are that the center of mass of an isolated system, initially at rest, remains at rest if no external forces act, and that electromagnetic radiation carries momentum E/c, as predicted by Maxwell's equations and experimentally established.

But how is this mass transfer physically realized? Is the front end of the tube really heavier after it absorbs the light? The answer is yes, because it's a bit hotter, which means its atoms are vibrating slightly faster—and faster moving objects have higher mass. (And there's another contribution we're about to discuss.)

Mass and Potential Energy

Suppose now at the far end of the tube we have a hydrogen atom at rest. As we shall see later, this atom is essentially a proton having an electron bound to it by electrostatic attraction. It is known that a flash of light with total energy 13.6eV is just enough to tear the electron away, so in the end the proton and electron are at rest far away from each other. The energy of the light was used up dragging the proton and electron apart—that is, it went into potential energy. (It should be mentioned that the electron also loses kinetic energy in this process, 13.6 ev is the net energy required to break up the atom.) Now, the light *is* absorbed by this process, so from our argument above the right hand end of the tube *must become heavier*. That is to say*, a proton at rest plus a (distant) electron at rest weigh more than a hydrogen atom* by E/c^2, with E equal to 13.6eV. Thus, Einstein's box forces us to conclude that increased *potential* energy in a system also entails the appropriate increase in mass.

It is interesting to consider the hydrogen atom dissociation in reverse—if a slow moving electron encounters an isolated proton, they may combine to form a hydrogen atom, emitting 13.6eV of electromagnetic radiation energy as they do so. Clearly, then, the hydrogen atom remaining has that much less energy than the initial proton + electron. The actual mass difference for hydrogen atoms is about one part in 10^8. This is typical of the energy radiated away in a violent chemical reaction—in fact, since most atoms are an order of magnitude or more heavier than hydrogen, a part in 10^9 or 10^{10} is more usual. However, things are very different in nuclear physics, where the forces are stronger so the binding is tighter. We shall discuss this later, but briefly mention an example: a hydrogen nucleus can combine with a lithium nucleus to give two helium nuclei, and the mass shed is 1/500 of the original. This reaction has been observed, and all the masses involved are measurable. The actual energy emitted is 17 MeV. This is the type of reaction that occurs in hydrogen bombs. Notice that the energy released is at least a million times more than the most violent chemical reaction.

As a final example, let's make a ballpark estimate of the change in mass of a million tons of TNT on exploding. The TNT molecule is about a hundred times heavier than the hydrogen atom, and gives off a few eV on burning. So the change in weight is of order 10^{-10} x10^6 tons, about a hundred grams. In a hydrogen bomb, this same mass to energy conversion would take about fifty kilograms of fuel.

Footnote: Einstein's Box is a Fake

Although Einstein's box argument is easy to understand, and gives the correct result, it is based on a physical fiction—the rigid box. If we had a rigid box, or even a rigid stick, all our clock synchronization problems would be over—we could start clocks at the two ends of the stick simultaneously by nudging the stick from one end, and, since it's a rigid stick, the other end would move instantaneously. Actually there are no such materials. All materials are held together by electromagnetic forces, and pushing one end causes a wave of compression to travel down the stick. The electrical forces between atoms adjust at the speed of light, but the overall wave travels far more slowly because each atom in the chain must accelerate for a while before it moves sufficiently to affect the next one measurably. So the light pulse will reach the other end of the box before it has begun to move! Nevertheless, the wobbling elastic box *does* have a net recoil momentum, which it does lose when the light hits the far end. So the basic point is still valid. French gives a legitimate derivation, replacing the box by its two (disconnected) ends, and finding the center of mass of this complete system, which of course remains at rest throughout the process.

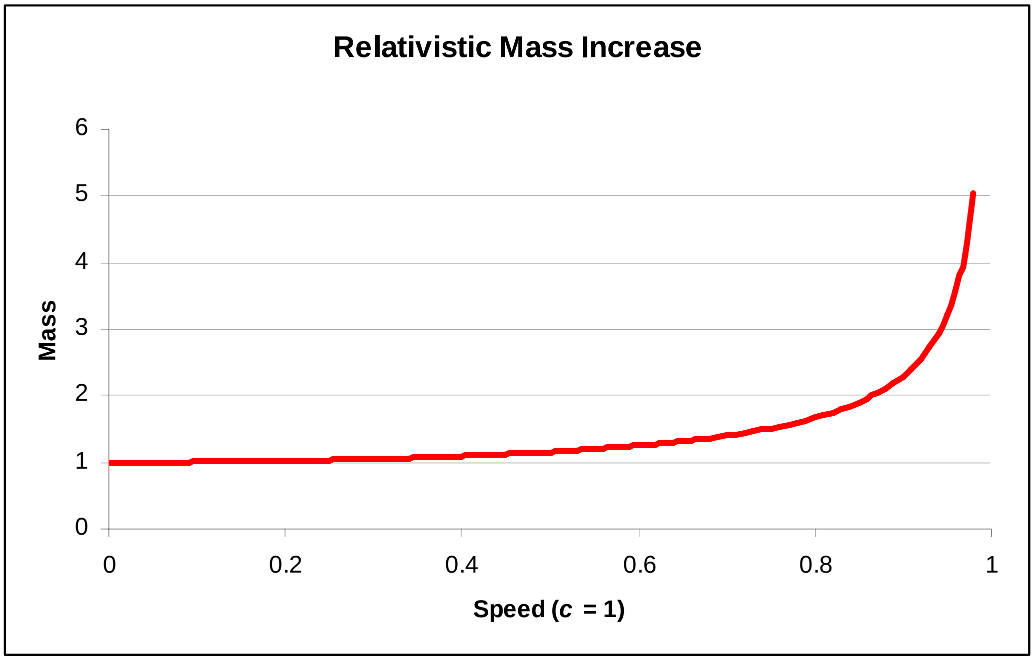
Energy and Momentum in Lorentz Transformations

How Does the Total Energy of a Particle Depend on Speed?

We have a formula for the total energy E = K.E. + rest energy,

$$E = mc^2 = \frac{m_0 c^2}{\sqrt{1 - v^2 / c^2}}$$

so we can see how total energy varies with speed.



The momentum varies with speed as

$$p = mv = \frac{m_0 v}{\sqrt{1 - v^2 / c^2}}.$$

How Does the Total Energy of a Particle Depend on Momentum?

It turns out to be useful to have a formula for E in terms of p.

Now

$$E^2 = m^2 c^4 = \frac{m_0^2 c^4}{1 - v^2 / c^2}$$

so

$$m^2 c^4 (1 - v^2 / c^2) = m_0^2 c^4$$
$$m^2 c^4 - m^2 v^2 c^2 = m_0^2 c^4$$
$$m^2 c^4 = E^2 = m_0^2 c^4 + m^2 c^2 v^2$$

hence using $p = mv$ we find

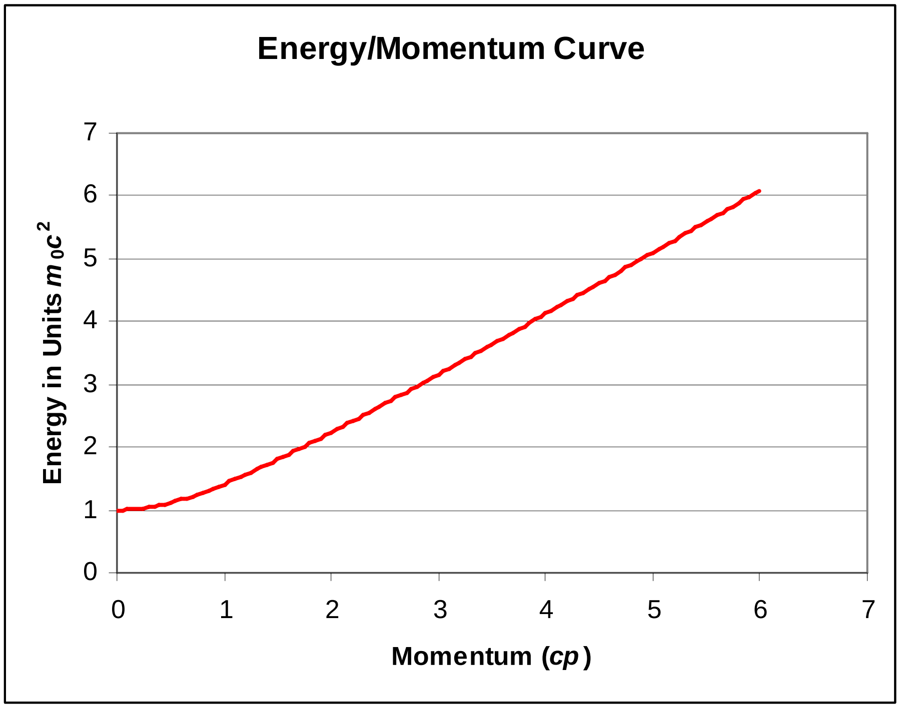
$$E = \sqrt{m_0^2 c^4 + c^2 p^2}.$$

If p is very small, this gives

$$E \approx m_0 c^2 + \frac{p^2}{2m_0},$$

the usual classical formula.

If p is very large, so $c^2 p^2 \gg m_0^2 c^4$, the approximate formula is $E = cp$.



The High Kinetic Energy Limit: Rest Mass Becomes Unimportant!

Notice that this high energy limit is just the energy-momentum relationship Maxwell found to be true for light, for all p. This could only be true for *all* p if $m_0^2 c^4 = 0$, that is, $m_0 = 0$.

Light is in fact composed of "photons"—particles having zero "rest mass", as we shall discuss later. The "rest mass" of a photon is meaningless, since they're never at rest—the energy of a photon

$$E = mc^2 = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}$$

is of the form 0/0, since $m_0 = 0$ and $v = c$, so "m" can still be nonzero. That is to say, the mass of a photon is really all K.E. mass.

For very fast electrons, such as those produced in high energy accelerators, the additional K.E. mass can be thousands of times the rest mass. For these particles, we can neglect the rest mass and take $E = cp$.

Transforming Energy and Momentum to a New Frame

We have shown

$$\vec{p} = m\vec{v} = \frac{m_0\vec{v}}{\sqrt{1-v^2/c^2}}$$

$$E = mc^2 = \sqrt{m_0^2c^4 + c^2\vec{p}^2}.$$

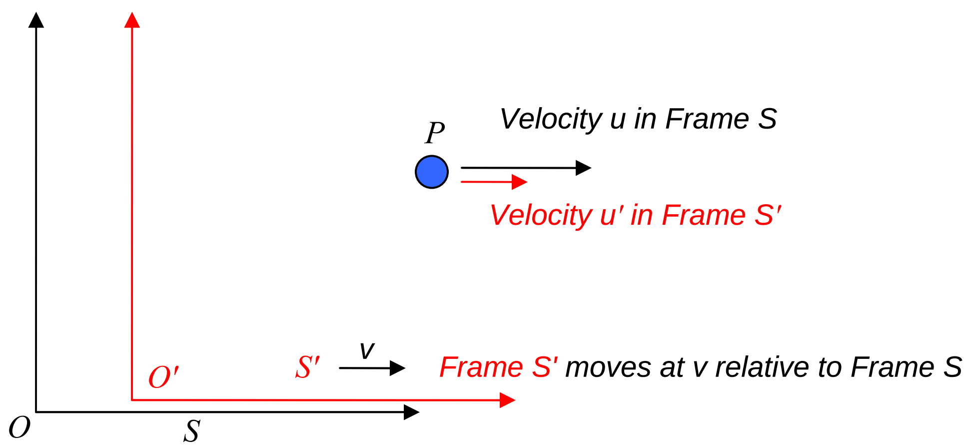
Notice we can write this last equation in the form

$$E^2 - c^2\vec{p}^2 = m_0^2c^4.$$

That is to say, $E^2 - c^2\vec{p}^2$ depends *only* on the rest mass of the particle and the speed of light. It does not depend on the velocity of the particle, so it must be the same—for a particular particle—in all inertial frames.

This is reminiscent of the invariance of $\vec{x}^2 - c^2t^2$, the interval squared between two events, under the Lorentz transformations. One might guess from this that the laws governing the transformation from E, p in one Lorentz frame to E', p' in another are similar to those for x,t. We can actually derive the laws for E, p to check this out.

As usual, we consider all velocities to be parallel to the x-axis.



We take the frame S' to be moving in the x-direction at speed v relative to S.

Consider a particle of mass m_0 (rest mass) moving at u' in the x' direction in frame S', and hence at u along x in S, where

$$u = \frac{u' + v}{1 + vu'/c^2}.$$

The energy and momentum in S' are

$$E' = \frac{m_0 c^2}{\sqrt{1 - u'^2/c^2}}, \quad p' = \frac{m_0 u'}{\sqrt{1 - u'^2/c^2}}$$

and in S:

$$E = \frac{m_0 c^2}{\sqrt{1 - u^2/c^2}}, \quad p = \frac{m_0 u}{\sqrt{1 - u^2/c^2}}.$$

Thus

$$E = \frac{m_0 c^2}{\sqrt{1 - \left(\frac{u' + v}{1 + vu'/c^2}\right)^2 / c^2}}$$

giving

$$E = \frac{m_0 c^2 \left(1 + vu'/c^2\right)}{\sqrt{\left(1 - v^2/c^2\right)\left(1 - u'^2/c^2\right)}}$$

from which it is easy to show that

$$E = \frac{1}{\sqrt{1 - v^2/c^2}} \left(E' + vp'\right).$$

Similarly, we can show that

$$p = \frac{p' + vE'/c^2}{\sqrt{1 - v^2/c^2}}.$$

These are the Lorentz transformations for energy and momentum of a particle—it is easy to check that

$$E^2 - c^2 p^2 = E'^2 - c^2 p'^2 = m_0^2 c^4.$$

Photon Energies in Different Frames

For a zero rest mass particle, such as a photon, $E = cp$, $E^2 - c^2p^2 = 0$ in all frames.

Thus

$$E = \frac{E' + vp'}{\sqrt{1 - v^2 / c^2}} = \frac{E' + vE'/c}{\sqrt{1 - v^2 / c^2}} = E'\sqrt{\frac{1 + v/c}{1 - v/c}}\ .$$

Since $E = cp$, $E' = cp'$ we also have

$$p = p'\sqrt{\frac{1 + v/c}{1 - v/c}}\ .$$

Notice that the ratios of photon *energies* in the two frames coincides with the ratio of photon *frequencies* found in the Doppler shift. As we shall see when we cover quantum mechanics, the photon energy is proportional to the frequency, so these two must of course transform in identical fashion. But it's interesting to see it come about this way. Needless to say, relativity gives us no clue on what the constant of proportionality (Planck's constant) is: it must be measured experimentally. But the *same* constant plays a role in all quantum phenomena, not just those concerned with photons.

Transforming Energy into Mass: Particle Creation

Pion Production

We have mentioned how, using a synchrocyclotron, it is possible to accelerate protons to relativistic speeds. The rest energy of a proton m_pc^2 is 938 MeV, using here the standard high energy physics energy unit: 1 MeV = 10^6 eV. The neutron is a bit heavier—m_nc^2 = 940 MeV. (The electron is 0.51 MeV). Thus to accelerate a proton to relativistic speeds implies giving it a K.E. of order 1,000 MeV, or 1 GeV.

The standard operating procedure of high energy physicists is to accelerate particles to relativistic speeds, then smash them into other particles to see what happens. For example, fast protons will be aimed at protons at rest (hydrogen atoms, in other words—the electron can be neglected). These proton-proton collisions take place inside some kind of detection apparatus, so the results can be observed. One widely-used detector is the bubble chamber: a transparent container filled with a superheated liquid. The electric field of a rapidly moving charged particle passing close to a molecule can dislodge an electron, so an energetic particle moving through the liquid leaves a trail of ionized molecules. These give centers about which bubbles can nucleate. The bubbles grow rapidly and provide a visible record of the particle's path.

What is actually observed in *p*-*p* scattering at relativistic energies is that often more particles come out than went in—particles called pions, π^+, π^0, π^- can be created. The π^0 is electrically neutral, the π^+ has exactly the same amount of charge as the proton. It is found experimentally that total electric charge is always conserved in collisions, no matter how many new particles are spawned, and total baryon number (protons + neutrons) is conserved.

Possible scenarios include:

$$p + p \to p + p + \pi^0$$

and

$$p + p \to p + n + \pi^+.$$

The neutral pion mass is 135 MeV, the charged pions have mass 140 MeV, where we follow standard high energy practice in calling mc^2 the "mass", since this is the energy equivalent, and hence the energy which, on creation of the particle in a collision, is taken from kinetic energy and stored in mass.

However, an incoming proton with 135 MeV of kinetic energy will not be able to create a neutral pion in a collision with a stationary proton. This is because the incoming proton also has momentum, and the collision conserves momentum, so some of the particles after the collision must have momentum and hence kinetic energy. The simplest way to figure out just how much energy the incoming proton needs to create a neutral pion is to go to the center of mass frame, where initially two protons are moving towards each other with equal and opposite velocities, there is no total momentum. Obviously, in this frame the least possible K.E. must be just enough to create the π^0 with all the final state particles (p,p,π^0) at rest. Thus if the relativistic mass of the incoming protons in the center of mass frame is m, the total energy

$$E = 2mc^2 = 2m_p c^2 + m_\pi c^2.$$

Putting in the proton and pion masses from above, and using $m = m_p / \sqrt{1-(v^2/c^2)}$, we find the two incoming protons must both be traveling at 0.36c.

Recall that this is the speed in the center of mass frame, and for practical purposes, like designing the accelerator, we need to know the energy necessary in the "lab" frame—that in which one of the protons is initially at rest. The two frames obviously have a relative speed of 0.36c, so to get the speed of the incoming proton in the lab frame we must add a velocity of 0.36c to one of 0.36c using the relativistic addition of velocities formula, which gives 0.64c. This implies the incoming proton has a relativistic mass of 1.3 times its rest mass, and thus a K.E. around 280 MeV.

Thus to create a pion of rest energy 135 MeV, it is necessary to give the incoming proton at least 290 MeV of kinetic energy. This is called the "threshold energy" for pion production. The "inefficiency" arises because momentum is also conserved, so there is still considerable K.E. in the final particles.

Antiproton Production

On raising the energy of the incoming proton further, more particles are produced, including the "antiproton"—a negatively charged heavy particle which will annihilate a proton in a flash of energy. It turns out experimentally that an antiproton can only be produced accompanied by a newly created proton,

$$p + p \to p + p + p + \overline{p}$$

Notice we could have conserved electric charge with less energy with the reaction

$$p + p \to p + p + \pi^+ + \overline{p}$$

but this doesn't happen—so charge conservation isn't the only constraint on which particles can be produced. In fact, what we are seeing here is experimental confirmation that the conservation of baryon number, which at the low energies previously discussed in the context of pion production just meant that the total number of protons plus neutrons stayed fixed, is generalized at high energies to include antiparticles with negative baryon number, -1 for the antiproton. Thus baryon number conservation becomes parallel to electric charge conservation, new particles can always be produced at high enough energies provided the total new charge and the total new baryon number are both zero. (Actually there are further conservation laws which become important when more exotic particles are produced, which we may discuss later.) We should emphasize again that these are *experimental* results gathered from examining millions of collisions between relativistic particles.

One of the first modern accelerators, built at Berkeley in the fifties, was designed specifically to produce the antiproton, so it was very important to calculate the antiproton production threshold correctly! This can be done by the same method we used above for pion production, but we use a different trick here which is often useful. We have shown that on transforming the energy and momentum of a particle from one frame to another

$$E^2 - c^2 p^2 = E'^2 - c^2 p'^2$$

Since the Lorentz equations are linear, if we have a system of particles with total energy E and total momentum p in one frame, E', p' in another, it must again be true that

$$E^2 - c^2 \vec{p}^2 = E'^2 - c^2 \vec{p}'^2.$$

We can use this invariance to get lab frame information from the center of mass frame. Noting that in the center of mass (CM) frame the momentum is zero, and in the lab frame the momentum is all in the incoming proton,

$$E_{cm}^2 = ((m_{in} + m_0)c^2)^2 - c^2 p_{in}^2$$

where here m_0 is the proton rest mass, m_{in} is the relativistic mass of the incoming proton. At the antiproton production threshold, $E_{cm} = 4m_0c^2$, so

$$16m_0^2c^4 = m_{in}^2c^4 + 2m_{in}c^2m_0c^2 + m_0^2c^4 - c^2p_{in}^2$$

and using

$$m_{in}^2c^4 - c^2p_{in}^2 = m_0^2c^4$$

we find

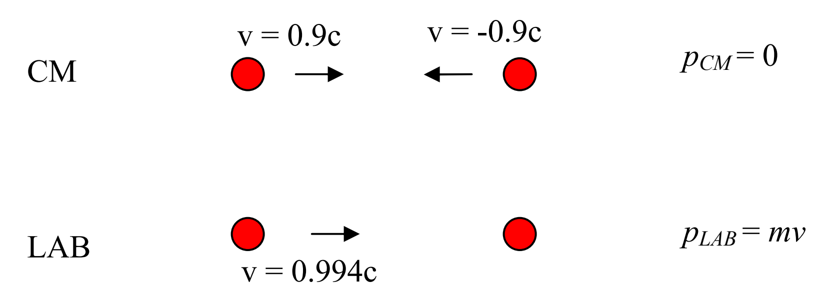
$$2(m_{in}c^2)(m_0c^2) + 2(m_0c^2)^2 = 16(m_0c^2)^2,$$

so

$$m_{in}c^2 = 7m_0c^2 .$$

Therefore to create two extra particles, with total rest energy $2m_0c^2$, it is necessary for the incoming proton to have a kinetic energy of $6m_0c^2$. The Berkeley Gevatron had design energy 6.2 GeV.

As we go to higher energies, this “inefficiency” gets worse—consider energies such that the kinetic energy >> rest energy, and assume the incoming particle and the target particle have the same rest mass, m_0, with the incoming particle having relativistic mass m.



The same collision as viewed in the CM and LAB frames.

Comparing the center of mass energy with the lab energy at these high energies,

$$\begin{aligned}
E_{LAB} &= \left(m + m_0\right)c^2, \\
E_{CM}^2 &= E_{LAB}^2 - p_{LAB}^2c^2 \\
&= m^2c^4 + 2mc^2m_0c^2 + m_0^2c^4 - p_{LAB}^2c^2 \\
&= 2m_0c^2(mc^2 + m_0c^2).
\end{aligned}$$

For $m >> m_0$,

$$E_{CM}^2 \approx 2m_0c^2mc^2 \approx 2m_0c^2.E_{LAB}$$

so

$$E_{CM} \approx \sqrt{2m_0c^2.E_{LAB}}\ ,$$

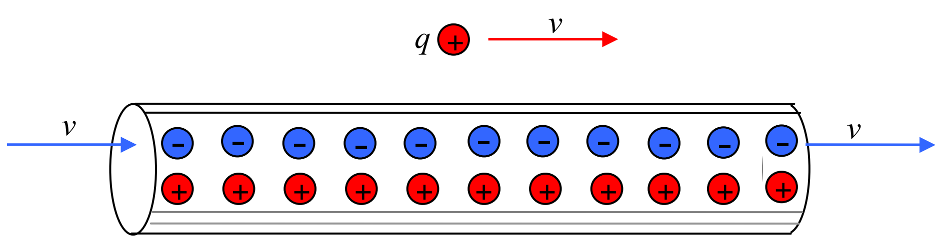
ultimately one must *quadruple* the lab energy to double the center of mass energy.

How Relativity Connects Electric and Magnetic Fields

A Magnetic Puzzle…

Suppose we have an infinitely long straight wire, having a charge density of electrons of $-\lambda$ coulombs per meter, all moving at speed v to the right (recall typical speeds are centimeters per minute) and a neutralizing fixed background of positive charge, also of course λ coulombs per meter. The current in the wire has magnitude $I = \lambda v$ (and actually is flowing to the left, since the moving electrons carry negative charge).

Suppose also that a positive charge q is outside the wire, a distance r from the axis, and this outside charge is moving at the same exact velocity as the electrons in the wire.



Positive charges (ions) are fixed: negative charges (electrons) drift at speed *v*.
Positive charge outside wire also moves at *v*.

Note: *positive and negative charges are of course uniformly distributed throughout the wire. They are shown separated here for clarity.*

What force does the positive charge q feel?

The wire is electrically neutral, since it contains equal densities of positive and negative charges, both uniformly distributed throughout the wire (the illustration above is of course schematic). So q feels no electrical force.

However, since q is moving, it will feel a magnetic force, $\vec{F}_{\text{mag}} = q\vec{v} \times \vec{B}$.

The magnetic field lines are vertical circles centered along the axis of the wire, the field being into the page where the charge is (remember the current is to the left) and of magnitude

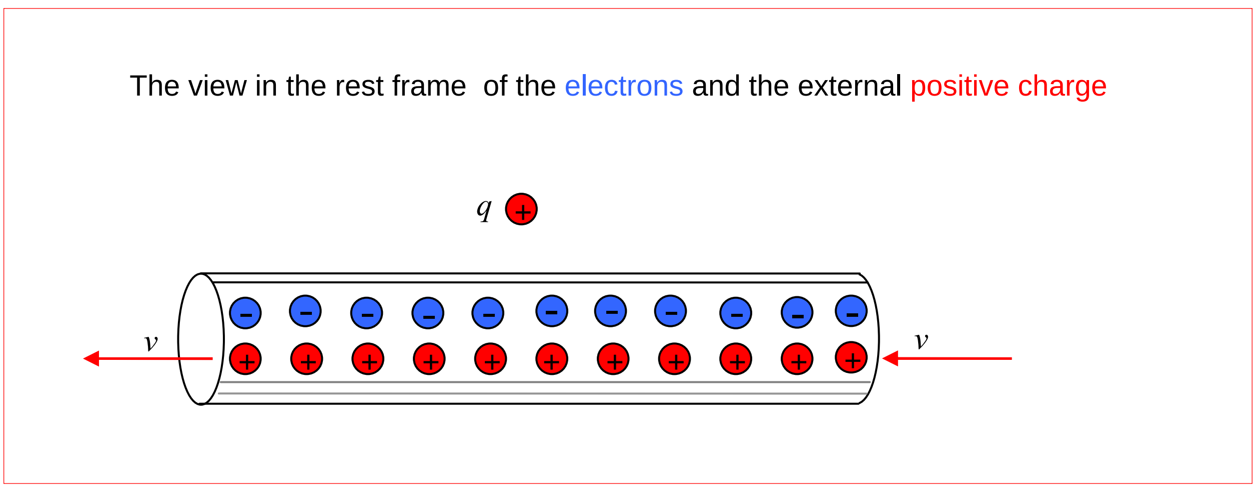
$$B = \frac{\mu_0 I}{2\pi r} = \frac{\mu_0 \lambda v}{2\pi r}$$

so the force on the charge is of magnitude

$$F = \frac{q\mu_0 \lambda v^2}{2\pi r}$$

and is directed away from the wire, so the charge will accelerate away from the wire.

Now let us examine the same physical system in the frame of reference in which the charge is initially at rest. In that frame, the electrons are also at rest, but the positive background charge is flowing at v:



What force does q feel in this frame?

Since q is at rest, it cannot feel a magnetic force: such forces depend linearly on speed!

Yet it looks as if it can't feel an electric force either, because the positive and negative charges in the wire have equal densities, right? This leads to the conclusion that q feels no force at all in this frame, so it won't accelerate away from the wire, as it did in the other frame. This is of course nonsense—so where did we go wrong?

… Solved Electrically!

The mistake was in ignoring the relativistic Fitzgerald-Lorentz contraction, even though the velocities involved are *millimeters per second*! In the frame in which the wire is at rest, the positive and negative charge densities exactly balance, otherwise there will be extra electrostatic fields that the electrons will quickly move to neutralize. However, this *necessarily* means that the densities cannot balance exactly in the frame in which the drifting electrons are at rest.

In the electrons' rest frame, the positive charges, which had density λ coulombs per meter in their rest frame, are moving at speed v, so relativistic contraction will increase their density to:

$$\frac{\lambda}{\sqrt{1-v^2/c^2}} \cong \lambda + \frac{\lambda v^2}{2c^2}.$$

On the other hand, in this electron frame the electrons are at rest, so their density is actually less than it was in the frame of the wire, it has decreased by $\lambda v^2/2c^2$.

The net effect is that the wire, electrically neutral in the lab frame, has a positive charge density $\lambda v^2/c^2$ in the frame of the moving electrons (and the outside charge). Recall Gauss's Law relating the total electric field flux through a surface to the enclosed charge,

$$\int \vec{E}.d\vec{A} = (\text{enclosed charge})/\varepsilon_0.$$

Exercise: use this to verify that the electric field from an infinite line of charge λ coulombs per meter has magnitude $\lambda/2\pi r\varepsilon_0$ at distance r from the line charge.

From the exercise result, the electrostatic force on the charge q is:

$$F = qE = q\lambda \frac{v^2}{c^2}\frac{1}{2\pi r\varepsilon_0} = \frac{q\lambda v^2}{2\pi r}\frac{1}{\varepsilon_0 c^2} = \frac{q\lambda v^2}{2\pi r}\mu_0$$

where in the last step we used $1/c^2 = \mu_0\varepsilon_0$.

This purely electrical force is identical in magnitude to the purely magnetic force in the other frame! So observers in the two frames *will* agree on the rate at which the particle accelerates away from the wire, but one will call the accelerating force magnetic, the other electric. We are forced to the conclusion that whether a particular force on an actual particle is magnetic or electric, or some mixture of both, depends on the frame of reference—so the distinction is rather artificial.

A much more complete analysis of the fields from moving charges can be found in Purcell, Electricity and Magnetism, McGraw Hill, 1985. We have chosen the simplest *possible example to illustrate the basic point that electric and magnetic fields are frame-dependent concepts.*

Remarks on General Relativity

Einstein's Parable

In Einstein's little book *Relativity: the Special and the General Theory*, he introduces general relativity with a parable. He imagines going into deep space, far away from gravitational fields, where any body moving at steady speed in a straight line will continue in that state for a very long time. He imagines building a space station out there - in his words, "a spacious chest resembling a room with an observer inside who is equipped with apparatus." Einstein points out that there will be no gravity, the observer will tend to float around inside the room.

But now a rope is attached to a hook in the middle of the lid of this "chest" and an unspecified "being" pulls on the rope with a constant force. The chest and its contents, including the observer, accelerate "upwards" at a constant rate. How does all this look to the man in the room? He finds himself moving towards what is now the "floor" and needs to use his leg muscles to stand. If he releases anything, it accelerates towards the floor, and in fact all bodies accelerate at the same rate. If he were a normal human being, he would assume the room to be in a gravitational field, and might wonder why the room itself didn't fall. Just then he would discover the hook and rope, and conclude that the room was suspended by the rope.

Einstein asks: should we just smile at this misguided soul? His answer is no - the observer in the chest's point of view is just as valid as an outsider's. In other words, *being inside* the (from an outside perspective) *uniformly accelerating room is physically equivalent to being in a uniform gravitational field*. This is the basic postulate of general relativity. Special relativity said that all inertial frames were equivalent. General relativity extends this to accelerating frames, and states their equivalence to frames in which there is a gravitational field. This is called the *Equivalence Principle*.

The acceleration could also be used to cancel an existing gravitational field—for example, inside a freely falling elevator passengers are weightless, conditions are equivalent to those in the unaccelerated space station in outer space.

It is important to realize that this equivalence between a gravitational field and acceleration is only possible because the gravitational mass is exactly equal to the inertial mass. There is no way to cancel out electric fields, for example, by going to an accelerated frame, since many different charge to mass ratios are possible.

As physics has developed, the concept of fields has been very valuable in understanding how bodies interact with each other. We visualize the electric field lines coming out from a charge, and know that something is there in the space around the charge which exerts a force on another charge coming into the neighborhood. We can even compute the energy density stored in the electric field, locally proportional to the square of the electric field intensity. It is tempting to think that the gravitational field is quite similar—after all, it's another inverse square field. Evidently, though, this is not the case. If by going to an accelerated frame the gravitational field can be made to vanish, at least locally, it cannot be that it stores energy in a simply defined local way like the electric field.

We should emphasize that going to an accelerating frame can only cancel a *constant* gravitational field, of course, so there is no accelerating frame in which the whole gravitational field of, say, a massive body is zero, since the field necessarily points in different directions in different regions of the space surrounding the body.

Some Consequences of the Equivalence Principle

Consider a freely falling elevator near the surface of the earth, and suppose a laser fixed in one wall of the elevator sends a pulse of light horizontally across to the corresponding point on the opposite wall of the elevator. Inside the elevator, where there are no fields present, the environment is that of an inertial frame, and the light will certainly be observed to proceed directly across the elevator. Imagine now that the elevator has windows, and an outsider at rest relative to the earth observes the light. As the light crosses the elevator, the elevator is of course accelerating downwards at g, so since the flash of light will hit the opposite elevator wall at precisely the height relative to the elevator at which it began, the outside observer will conclude that the flash of light also accelerates downwards at g. In fact, the light could have been emitted at the instant the elevator was released from rest, so we must conclude that light falls in an initially parabolic path in a constant gravitational field. Of course, the light is traveling very fast, so the curvature of the path is small! Nevertheless, *the Equivalence Principle forces us to the conclusion that the path of a light beam is bent by a gravitational field*.

The curvature of the path of light in a gravitational field was first detected in 1919, by observing stars very near to the sun during a solar eclipse. The deflection for stars observed very close to the sun was 1.7 seconds of arc, which meant measuring image positions on a photograph to an accuracy of hundredths of a millimeter, quite an achievement at the time.

One might conclude from the brief discussion above that a light beam in a gravitational field follows the same path a Newtonian particle would if it moved at the speed of light. This is true in the limit of small deviations from a straight line in a constant field, but is not true even for small deviations for a spatially varying field, such as the field near the sun the starlight travels through in the eclipse experiment mentioned above. We could try to construct the path by having the light pass through a series of freely falling (fireproof!) elevators, all falling towards the center of the sun, but then the elevators are accelerating relative to each other (since they are all falling along *radii*), and matching up the path of the light beam through the series is tricky. If it is done correctly (as Einstein did) it turns out that the angle the light beam is bent through is twice that predicted by a naïve Newtonian theory.

What happens if we shine the pulse of light vertically *down* inside a freely falling elevator, from a laser in the center of the ceiling to a point in the center of the floor? Let us suppose the flash of light leaves the ceiling at the instant the elevator is released into free fall. If the elevator has height h, it takes time h/c to reach the floor. This means the floor is moving downwards at speed gh/c when the light hits.

Question: Will an observer on the floor of the elevator see the light as Doppler shifted?

The answer has to be no, because inside the elevator, by the Equivalence Principle, conditions are identical to those in an inertial frame with no fields present. There is nothing to change the frequency of the light. This implies, however, that to an outside observer, stationary in the earth's gravitational field, the frequency of the light *will* change. This is because he will agree with the elevator observer on what was the initial frequency f of the light as it left the laser in the ceiling (the elevator was at rest relative to the earth at that moment) so if the elevator operator maintains the light had the same frequency f as it hit the elevator floor, which is moving at gh/c relative to the earth at that instant, the earth observer will say the light has frequency $f(1 + v/c) = f(1+gh/c^2)$, using the Doppler formula for very low speeds.

We conclude from this that light shining downwards in a gravitational field is shifted to a higher frequency. Putting the laser in the elevator floor, it is clear that light shining upwards in a gravitational field is red-shifted to lower frequency. Einstein suggested that this prediction could be checked by looking at characteristic spectral lines of atoms near the surfaces of very dense stars, which should be red-shifted compared with the same atoms observed on earth, and this was confirmed. This has since been observed much more accurately. An amusing consequence, since the atomic oscillations which emit the radiation are after all just accurate clocks, is that *time passes at different rates at different altitudes*. The US atomic standard clock, kept at 5400 feet in Boulder, gains 5 microseconds per year over an identical clock almost at sea level in the Royal Observatory at Greenwich, England. Both clocks are accurate to one microsecond per year. This means you would age more slowly if you lived on the surface of a planet with a large gravitational field. Of course, it might not be very comfortable.

General Relativity and the Global Positioning System

Despite what you might suspect, the fact that time passes at different rates at different altitudes has significant practical consequences. An important *everyday* application of general relativity is the Global Positioning System. A GPS unit finds out where it is by detecting signals sent from orbiting satellites at precisely timed intervals. If all the satellites emit signals simultaneously, and the GPS unit detects signals from four different satellites, there will be three relative time delays between the signals it detects. The signals themselves are encoded to give the GPS unit the precise position of the satellite they came from at the time of transmission. With this information, the GPS unit can use the speed of light to translate the detected time delays into distances, and therefore compute its own position on earth by triangulation.

But this has to be done very precisely! Bearing in mind that the speed of light is about one foot per nanosecond, an error of 100 nanoseconds or so could, for example, put an airplane off the runway in a blind landing. This means the clocks in the satellites timing when the signals are sent out must certainly be accurate to 100 nanoseconds a day. That is one part in 10^{12}. It is easy to check that both the special relativistic time dilation correction from the speed of the satellite, and the general relativistic gravitational potential correction are much greater than that, so the clocks in the satellites must be corrected appropriately. (The satellites go around the earth once every twelve hours, which puts them at a distance of about four earth radii. The calculations of time dilation from the speed of the satellite, and the clock rate change from the gravitational potential, are left as exercises for the student.) For more details, see the lecture by Neil Ashby here.

In fact, Ashby reports that when the first Cesium clock was put in orbit in 1977, those involved were sufficiently skeptical of general relativity that the clock was not corrected for the gravitational redshift effect. But—just in case Einstein turned out to be right—the satellite was equipped with a synthesizer that could be switched on if necessary to add the appropriate relativistic corrections. After letting the clock run for three weeks with the synthesizer turned off, it was found to differ from an identical clock at ground level by precisely the amount predicted by special plus general relativity, limited only by the accuracy of the clock. This simple experiment verified the predicted gravitational redshift to about one percent accuracy! The synthesizer was turned on and left on.